

Distasteful but necessary

The public must be told that experiments on primates remain essential for progress in some areas of biomedicine. But the scientists involved should also lead the way in pressing for improvements in animal welfare.

The initial reaction of one British researcher to the news that *Nature* was planning an article examining the use of primates in biomedical experiments was telling. He expressed deep misgivings, arguing that the climate of opinion in Britain is so unfavourable that any publicity could only be damaging to researchers who, like him, work on monkeys.

When rational people abandon public debate, the outlook is bleak. But in a country where researchers live in fear of animal-rights terrorists, and where an application to build a primate research facility was turned down because the local authority feared the implications for public order of protests against its construction, such fatalism is perhaps understandable, although misguided.

As *Nature's* overview of the current practice of primate research reveals, there is a strong case to be made for its continuation (see page 684). If researchers were not able to infect rhesus macaques with viruses closely related to HIV, for instance, we could hold out little hope of developing a vaccine against AIDS.

At one level, considering whether a primate experiment is justified is simple. The potential benefits must be weighed against the suffering caused, bearing in mind that primates may have a greater capacity for suffering than other animals. But this deceptively simple equation is devilishly difficult to balance, particularly if the benefits are not obvious to non-specialists.

Many people reluctantly accept experiments on primates where there is a clear prospect of preventing or curing a deadly disease. But electrophysiological recordings from monkeys' brains, for example, attract more public resistance. In such cases, scientists and the bodies that represent them must do a much better job in explaining

how such fundamental research is, in the long run, likely to lead to improved treatments for human disease. They should also explain the measures they take to minimize animal suffering, making it harder for protesters to misrepresent researchers as uncaring monsters.

Scientists in countries where protests against primate research have not escalated to violence should take a lead. Some US researchers fear that the British experience might be exported — militant animal-rights activists are networking across the Atlantic. The time to inform public opinion is now, not when the threat seems more immediate, when the battle for hearts and minds will already be half lost.

Researchers should do more than simply defend the status quo. Their case will be more compelling if they are seen to lead the way in pressing for a reduction in the use of primates where possible, and for improvements in animal welfare. For the former, attention should focus on whether some of the drug toxicity tests currently conducted on monkeys need to be done. If alternatives are avoided simply because companies don't want to risk regulators rejecting data from a lab animal with which they are unfamiliar, things need to change.

To boost animal welfare, researchers should push for dedicated facilities supported with veterinary scientists and specialists in primate behaviour. The days of labs holding primates in conditions that fail to meet the animals' behavioural needs should be over. The facilities also need self-sufficient breeding colonies — it is unacceptable for science to contribute to the decline of wild populations.

Some recent developments, such as the Dutch government's investment in the Biomedical Primate Research Centre in Rijswijk, are encouraging. But more must be done to improve facilities for research that, although distasteful, remains essential. ■

Pay or train to publish?

China's authorities place too much emphasis on the former and too little on the latter.

Most of the internationally recognized research by Chinese scientists is done outside China. In response, the Chinese government and universities pay indigenous researchers bonuses for getting their papers into international journals. Such bonuses, which can amount to thousands of dollars, have provided motivation. But they reflect an unbalanced focus on ends while neglecting the means: doing creative science. Moreover, impatience to secure research positions or to win such bonuses are partly responsible for problems in China with plagiarism and other misconduct.

China's graduate education system needs serious attention. Students in China and Chinese researchers who have teaching experience abroad criticize university and graduate education in China for its focus on textbook material and memorizing facts. As is well recognized in China, such a system cannot develop the creativity needed to achieve the international recognition in science that the universities and government seek. Many of the most promising students, who are also aware of the problem, go abroad for education or postdoctoral training — often never to come back.

A biological research institute in Shanghai (see page 683) is trying

to solve this problem. Teachers at its first-year graduate-school courses ask students to look at the latest international research, and teach them to ask questions about it rather than merely memorize it. The courses provide not only the desire but also the tools to do creative science. The students are rising to the challenge.

Such courses could be established elsewhere in China, and not only in biological sciences. Finding researchers who are able and willing to teach the courses, which are taught in Chinese, will not be easy. The courses in Shanghai depend on Chinese researchers based at top research institutes or universities in the United States, or who have had years of training outside China. With a budget that only covers their expenses, the researchers are working for free. The number of Chinese researchers worldwide who could teach the courses is increasing, but the funds are currently not in place.

The Shanghai scheme might motivate other university professors in China to look at different teaching methods and try to improve their own. Government support for such initiatives is needed. Researchers chasing bonuses can stimulate some quick results, but revamping graduate education would be a better bet for China's future. ■





Brain power
France trials
electrode treatment
for obsessive disorder
p677



Going short
UK science centres
seek new ways to
make money
p678



Peso peril
Argentina struggles
to fund cosmic-ray
observatory
p679



Joint effort
Nanotechnology
urged to forge
links with biology
p681

Homeland security plan sparks fears for fight against bioterror

Geoff Brumfiel, Washington

President George W. Bush's proposal for a new Department of Homeland Security — unveiled on 6 June — includes the reorganization of large portions of the US government's research portfolio. And scientific leaders are already fretting that the plan has the potential to disrupt important research in medicine, agriculture and nuclear weapons.

The proposed department would absorb scientific programmes worth about \$3.5 billion annually from the National Institutes of Health (NIH), the Department of Energy and other agencies to support its home defence mission. According to White House documents, the transfers would include \$1.75 billion that is currently set to be spent by the NIH next year on civilian biodefence, as well as the entire \$1.2-billion budget of the Lawrence Livermore National Laboratory (LLNL) in California, one of the United States' nuclear-weapons laboratories.

Mission statement

"I think there's a tremendous advantage to be gained by having an organization with a primary mission of homeland security, and that goes for science as well," says John Marburger, President Bush's science adviser.

Leaders of the programmes to be transferred, including Anthony Fauci, director of the National Institute for Allergies and Infectious Diseases (NIAID), which currently manages most of the NIH's biodefence funding, and Bruce Tarter, director of the LLNL, say they were not consulted about the proposed transfers. And scientific societies and other observers have questioned the usefulness and feasibility of the transfers. Congress is expected to probe the reform plan carefully and pass legislation to implement it, possibly after extensive modification, by the autumn.

The new department's research directorate would absorb the NIH's yearly biodefence budget of \$1.75 billion, but officials in the Bush administration suggest that the new department will send at least some of the money straight back to the NIH for disbursement to researchers. "There's been some talk about contracting or leasing functions



Unwieldy task: President Bush plans to bring US science to bear in the fight against terrorism.

from the NIH and not transferring them completely," says Tommy Thompson, the Health and Human Services Secretary.

But such arrangements are notoriously difficult to sustain between US government agencies. "From an operational point of view, we believe this plan is fraught with dangers," says Janet Shoemaker, head of public affairs at the American Society of Microbiology. She believes that separating biodefence research from the NIH's work on infectious diseases could weaken both programmes.

If biodefence researchers report to a new department that is less orientated towards science than the NIH, the quality of their research will suffer, critics suggest. They also question the wisdom of separating the study of diseases manipulated by humans from those that occur naturally.

The plan to transfer responsibility for LLNL may be opposed by scientists and congressmen who want the laboratory to concentrate on its nuclear-weapons work. "You can't expect a national lab to turn on a dime," says Chris Paine of the Natural

Resources Defense Council, an environmental group that monitors the laboratory. Only 18 months ago, the laboratory was placed under the control of the Department of Energy's new National Nuclear Security Administration in a bid to simplify and clarify responsibility for its management in the wake of the Wen Ho Lee spying scandal.

Security concern

"The idea is to have a national-laboratory-type enterprise whose highest priority is homeland security," Marburger explains. LLNL would receive funding for research and development related to homeland security, he adds, and the money could be spent on research either within the laboratory or outside it.

"It's a very good idea for the Department of Homeland Security to have a lead laboratory that it can turn to," says former LLNL director Michael May.

"There is already a substantial capability at Livermore," agrees John Holdren, director of the Science, Technology and Public Policy Program at Harvard University. Holdren adds that a new mission for the lab would cause few problems, provided that nuclear-weapons research is allowed to continue.

The plan also calls for the transfer into the new department of the Department of Agriculture's Animal and Plant Health Inspection Service. Tony Mazzaschi, vice-president for health research at the Association of American Medical Colleges, says that the cash-strapped service might benefit, but that it is hard to see how its main task — inspecting farms and animal-research facilities — will fit into the home security mission.

And despite Bush's pledge to implement the 6 June plan in full, most details remain unresolved, and will only take firm shape as Congress moves to craft its enabling legislation. "It is one of those situations where sweeping decisions are made and the details worked out later," says Marburger. While this is happening, researchers in the affected areas will look on with some apprehension. "Our initial reaction to this," says Shoemaker, "is one of great anxiety."

Additional reporting by Erika Check and Virginia Gewin.

US set to rejoin international fusion project

Geoff Brumfiel, Princeton

Government officials have indicated that the United States is ready to rejoin ITER, the international project to build an experimental magnetic fusion reactor.

Speaking at celebrations to mark Princeton Plasma Physics Laboratory's fiftieth birthday on 6 June, energy under-secretary Robert Card said that a decision on whether to return to ITER will come "within weeks rather than months".

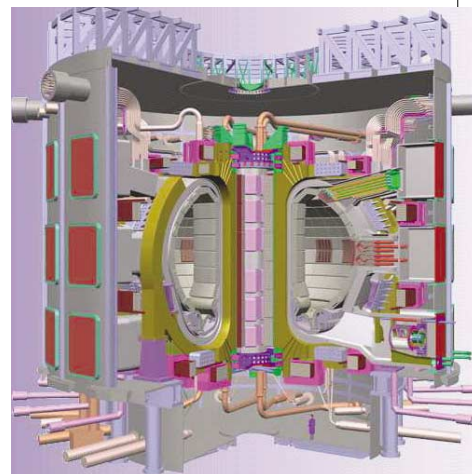
"We've more or less decided to re-enter talks" with ITER's current participants, says presidential science adviser John Marburger. But, he adds, the United States needs to have a firmer understanding of the project's costs and implications before making a firm financial commitment to it.

The extent of the United States' planned involvement remains an open question, but Marburger says that the country will not offer to pay the \$1 billion or more needed to

host the experimental reactor. "We're not interested in having ITER in the United States," he says. Observers say that US participation would probably involve a commitment of about \$500 million over 10 years.

The United States and the Soviet Union started ITER, then known as the International Thermonuclear Experimental Reactor, as part of détente in 1988. But just over a decade later, when the engineering phase yielded a design that could have cost up to \$10 billion to build, Congress instructed the energy department to withdraw from the project. This withdrawal was accompanied by sharp cuts in the domestic fusion budget, leaving US researchers with no prospect of building a large fusion experiment of their own.

Subsequently, the remaining ITER participants — Russia, Canada, Japan and the European Union — have developed a slimmed-down version of the design, which is now expected to cost about \$4 billion to



The United States could yet take part in ITER — but it has ruled out hosting the reactor itself.

ITER

build. At a meeting in Cadarache, France, on 6 June, the project members formally announced four potential sites for the reactor — two in Europe, one in Canada and one in Japan.

In the past year, both members of Congress and the Bush administration have steadily warmed to the idea of once again participating in ITER (see *Nature* **415**, 247–248; 2002).

US researchers reacted with a mixture of caution and optimism to the prospect of the United States rejoining ITER. "I think this is an opportunity for the United States to re-engage with the international programme," says Rob Goldston, director of the Princeton Plasma Physics Laboratory in New Jersey. "But I think it is critical that the United States strengthens its domestic programme as part of such an initiative."

The domestic US fusion-research programme "is still pretty fragile," says Gerry Navratil, a physicist at Columbia University. He is worried that its management will be distracted by negotiations over ITER, and even questions whether US scientists will gain full access to data from the international experiment. He wants the United States to establish a firm withdrawal date should the negotiations prove not to be fruitful, and to continue working on its own, more modest, plasma experiment.

As ITER negotiators begin discussing their choice of site, US researchers will debate the relative merits of ITER and a domestic experiment next month at a conference in Snowmass, Colorado. Goldston hopes that the negotiations abroad and discussion at home will bring the United States closer to the ultimate goal of building a viable fusion reactor. "Between vision and reality a number of steps remain," he says, "but I think it's possible."

♦ www.iter.org

Detectors licked by gummy fingers

David Cyranoski, Tokyo

Fingerprint-identification equipment can readily be fooled by a piece of gelatin, according to a cursory study undertaken by a Japanese mathematician.

Tsutomu Matsumoto of Yokohama National University says that his findings could undermine the extravagant claims being made for biometry, which uses inherent human traits such as fingerprints to automatically identify individuals.

In Matsumoto's informal study, he made plastic moulds of the subject's fingers, poured gelatin into the moulds, and let the thin 'gummy fingers' harden. Hundreds of trials by 5 people on 11 devices showed that a person wearing the gummy finger could pass for the subject nearly every time, he says.

Even more alarmingly, Matsumoto was able to create an effective fake fingerprint from a piece of glass that the subject had touched. He used a digital camera to photograph an enhanced copy of the mark, etched it on a printed circuit board, and used this to produce a fake finger made of gelatin.

"What's really scary is that anyone could do this," says Matsumoto. "The technology is all on the Internet and can be done cheaply at home."

He emphasizes that he is only dabbling in this research "for fun" while he pursues his real work on the application of cryptography and mathematics to information security.

Spokespeople at Sony and Fujitsu, which make automated fingerprint readers, declined to deny that their devices could be



Keep 'em peeled: fake fingerprints can readily be made using gelatin and a simple mould.

deceived by Matsumoto's gummy finger. But Sony and NEC, a third manufacturer, say that they are currently developing better systems. The new NEC reader, for example, will use light from the side that diffuses through the subject's finger, instead of just reflecting off the fingerprint.

Matsumoto is in no doubt that more sophisticated fingerprint-readers could pass his existing test — but he questions the philosophy of relying on biometry for security. "A password can be changed if it gets leaked," says Matsumoto, "but once someone has a copy of your fingerprint or your DNA, they will always be able to use it."

France wires up to treat obsessive disorder

Sally Goodman, Paris

France's ethics commission has given the go-ahead for clinical trials of neurosurgery on patients with obsessive-compulsive disorder (OCD), a debilitating psychiatric condition.

The ruling by the National Consultative Committee on Ethics will allow researchers to conduct carefully controlled trials using neurostimulation—a technique that involves stimulating a targeted area of the brain using implanted electrodes connected to a battery.

Neurosurgery was virtually abandoned as a treatment of psychiatric illness throughout the world after public revulsion over the abuse of lobotomies half a century ago. However, progress in neuroscience has led to renewed interest in such interventions, raising some thorny ethical questions.

Neurostimulation has the advantage of being reversible and, because the stimulator can be switched on and off and the electrodes removed, blind control trials can be carried out on consenting patients. OCD is a leading contender for such trials because it is well understood and therefore easy to target in the brain. What's more, its sufferers, unlike most other psychiatric patients, can often give informed consent to surgery. In both the United States and Europe, ethics committees decide whether the patients can give consent on a case-by-case basis.

Last October, Alim-Louis Benabid, a neurosurgeon at the Joseph Fourier University in Grenoble, asked the French commission to consider the ethics of using neurostimulation on OCD patients. The technique has been used to treat movement disorder in Parkinson's disease, but it has been tried on only about 20 OCD patients in Belgium, Sweden and the United States in the past five years.

Experimental surgical techniques are not constrained by France's bioethics laws, but Benabid says that he wanted an opinion from the ethics commission before starting clinical trials. The commission decided in favour of using neurostimulation, but said that strict experimental conditions should apply, including the setting up of a committee to oversee the choice of patients, their consent, and the evaluation of results.

The commission added that the experiment should involve only the most severe cases of OCD. Sufferers perform rituals, such as washing their hands, hundreds of times a day. Around 1–2% of people are thought to be affected, with some 10% of sufferers showing severely debilitating symptoms.

Bart Nuttin, a neurosurgeon from the Catholic University of Leuven in Belgium, told the commission that he was working with others to draft a scientific and ethical framework for using the technique before it



Tried and tested: neurostimulation has been used to treat patients with Parkinson's disease.

becomes widely established.

"We were very concerned about this technique falling into the wrong hands," Nuttin says, adding that the journal *Neurosurgery* has accepted the group's proposed guidelines for publication later this year.

However, progress in functional brain imaging is also tempting researchers to try to use neurostimulation to treat other psychiatric illnesses, such as depression, where such issues as the identification of the organic basis of the disorder and informed consent are even more complex. ■

C. POUEDRAS/EURELIOS/SPL

Statistical error leaves pollution data up in the air

Jonathan Knight, San Francisco

An off-the-shelf statistics package has tripped up pollution researchers in North America and Europe who are studying the effects of airborne soot on human health.

A default setting that produced erroneous results went unchecked for years, despite significant statistical expertise in all of the groups. "It was already such standard software when we started using it, I didn't

even question it," says Francesca Dominici, a public-health researcher at Johns Hopkins University in Baltimore, Maryland.

On 4 June, Dominici posted revised figures on her website after discovering that the error had doubled her group's estimate of the risk to health posed by particulates in the air. Two other groups that used the same tool, one in Canada and one in Greece, are now redoing their calculations.

The groups were looking for correlations between death rates and particulates in the air, which come mainly from diesel engines and power plants. Their data on air quality, hospitalizations and deaths from dozens of cities cover a seven-year period up to 1994.

Death rates vary throughout the year because of such factors as changes in temperature and disease outbreaks. To tease out the effects of particulates, the groups used a statistics program known as S-Plus.

S-Plus searches for correlations using an iterative process in which confounding effects are gradually peeled away. The default parameter in question determined how many times the procedure would iterate before stopping to produce a final result.

"For most applications the value is perfectly fine," says David Smith, product manager of Seattle-based Insightful, which sells S-Plus. Smith says that the Hopkins case was exceptional, but that users should always check whether changing the parameter affects the outcome, and adjust it if necessary. Smith says that Insightful will tighten the default value of the parameter—slowing the programme slightly—on future releases of S-Plus.

Richard Burnett, a statistician with Health Canada in Ottawa, which is conducting a similar study, says that his group will probably revise its estimates of the impact of airborne soot on mortality downwards by 20–50%. The findings of a study run by a group at the University of Athens may also have to be adjusted, he says.

The health risk posed by particulates is a source of fierce environmental controversy in the United States, and the Bush administration is considering rules to restrict emissions. Opponents of tighter rules are likely to seize on the revisions as evidence that the research linking soot in the air to harmful effects on health is not to be trusted. ■



Clouding the issue: particulates are seen as a health risk, but how reliable are the data?

D. MCNEW/NEWSMAKERS

Science centres struggle as funds run out

David Adam, London

As the Millennium Dome in London stands empty and forlorn by the River Thames, doubts are growing about the future of a network of science centres set up in Britain as another part of the millennium celebrations.

A dozen or so centres — conceived to boost the public understanding of science, educate children and help to revitalize depressed urban areas — have been built over the past few years largely with £250 million (US\$360 million) raised from Britain's national lottery. But with no more lottery funds available to subsidize them, many of them are already finding the going tough.

Typical difficulties are being confronted this month by the Glasgow Science Centre, which opened last summer in a striking

titanium-clad building on the banks of the River Clyde. Already heavily in debt because of delays, cost overruns and technical faults, the centre's trustees agreed at a meeting last week to take £2 million from a £7-million endowment fund, originally set aside to buy new exhibits, to help balance the books.

The move buys the centre time, but some observers are warning that if public money is not forthcoming, Glasgow and many of the other centres face cutbacks or even closure.

"Without support from government, the future of science centres that do not have income streams other than visitors' spending is bleak," says Linda Conlon, chief executive of the International Centre for Life, a science centre in Newcastle. Her centre hosts exhibitions and even theatre productions on issues

across the life sciences. But it also runs a conference and banqueting business, rents space to local businesses, and hosts the University of Newcastle's Institute of Human Genetics.

"We knew that if we delivered a science centre on its own with no other revenue streams we would be dead within 12 months," Conlon says. She estimates that only about two-thirds of most centres' costs can be met by visitors' admission fees and the money spent in cafeterias and bookshops.

John Durant, chief executive of a centre called at-Bristol, is critical of the model used to set up the centres by the Millennium Commission — the body that allocated lottery money to the projects — in which no money was set aside for future operating costs. "It was always a complete nonsense and they were told at the time it was nonsense," he says. Durant warns that many centres could soon be forced to replace educational activities with more commercial alternatives.

Mike O'Connor, director of the Millennium Commission, says that bidders were told that there would be no long-term support and that the commission nevertheless "received applications from people who said that their centres would get sufficient visitor numbers and enough income to be viable," he argues.

Melanie Quin, executive director of Ecsite-uk, an umbrella network for the UK centres, admits that the visitor numbers predicted by some of the centres' plans were "wishful thinking". Quin suggests that the government could step in and pay the centres to play a more direct role in school education. ■



Science centres such as those at Newcastle (inset) and Glasgow can't survive on visitors' money alone.

Environment institute slammed over 'selective' results

Johanna Schwarz and Alison Abbott, Munich

Germany's science council, the Wissenschaftsrat, has issued a scathing evaluation of a research institute that has played an important role in influencing the country's energy and environmental policies.

The assessment, which was published last month, says that the Wuppertal Institute near Dusseldorf should redefine its mission and improve the quality of its scientific research — or else no longer receive public money.

The Wuppertal Institute was founded in 1991 by the state of Nordrhein Westfalen — which supplies most of its funds — to bridge the gap between basic science and political strategy on environmental issues. But some scientists claim that its research findings have been driven more by ideology than by science.

The Wissenschaftsrat's report, which was requested by the state government, says that some of the institute's departments work well. But the climate department, it says,

"follows a one-sided concept, thereby weakening the scientific basis of its political advisory role". It adds that the department gives the impression of being selective in its consideration of research results.

The institute should concentrate on fewer projects, interact more with outside scientists, and publish more of its work in peer-reviewed scientific journals, it recommends.

The Wuppertal Institute says that it will try to comply with the recommendations, but has branded parts of the evaluation as "unfair". It argues that the Wissenschaftsrat has ignored its recent efforts to improve its research and failed to take into account its mission to provide policy advice as well as scientific results.

But Reinhard Hüttel, a soil scientist at the Brandenburg University of Technology in Cottbus who headed the evaluation committee, says that "every important detail of the institute" was considered.



Some scientists have sympathy with the institute's predicament. The institute "was founded to foster public debate on environmental policy, not to produce top-notch research", says Carlo Jäger, a social scientist at the Potsdam Institute for Climate Impact Research. ■

Argentine crisis rattles cosmic-ray hunters

Carol Marzuola

Astrophysicists awaiting the Pierre Auger Observatory's high-energy cosmic-ray data are keeping a nervous eye on the political and economic crisis in Argentina, where the Southern Hemisphere arm of the observatory is due to be completed by 2005.

Two hundred and fifty researchers from 19 countries are involved in the US\$50-million project to build the world's largest cosmic-ray observatory. It will study cosmic radiation — the flow of charged particles from space that streams through the Earth's atmosphere — and, in particular, the rare and mysterious particles of very high energies.

The southern part of the observatory, in Argentina's Mendoza province, is being built first. The northern arm, to be located in the United States, will come later. Argentina had pledged \$15 million to the project, but since December 2001, when the Argentinian peso was devalued, the country has plunged deeper into economic and political crisis (see *Nature* **415**, 104; 2002) and is unlikely to fulfil all of its obligations.

High-energy cosmic rays intrigue physicists because they cannot identify astrophysical objects or events, such as colliding galaxies, that are sufficiently violent to explain the rays' origins. But high-energy particles of 10^{20} electron volts or more hit the Earth's atmosphere only rarely — about once per km^2 per century.

The Auger observatory will use two sets of detectors to find these cosmic rays. At each site, 1,600 surface detectors — large above-ground water tanks equipped with photomultipliers — will identify secondary cosmic-ray showers produced in the atmosphere by incoming rays. Additionally, 24 special telescopes will detect fluorescence produced by the particles as they pass through the atmosphere.

A prototype array of 30 detectors and two telescopes was completed last year in Argentina and has already produced useful data, says James Cronin of the University of Chicago, a joint leader of the Auger project.

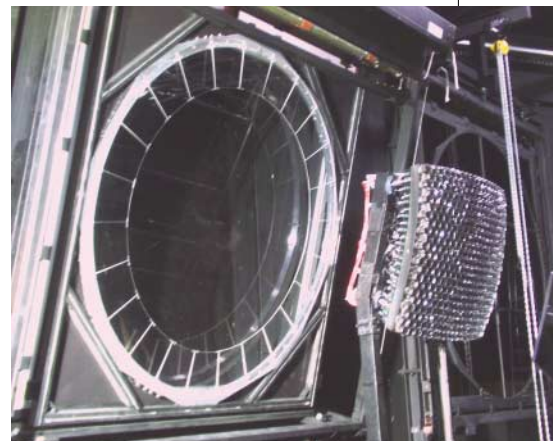
Alberto Etchegoyen, an official at Argentina's atomic-energy commission and the project's southern spokesman, says that the transitional Argentinian government has no clear long-term plans for science or technology, and admits that the project may have to find some of the promised funding from other sources. Project supporters in Argentina and Brazil, for example, are jointly seeking a \$3-million grant from an international development bank fund for Brazil to enable them to participate.

Optimists also point out that the devaluation of the peso will make the dollars and euros of the project's other contributing nations go further in Argentina. Paul



Mantsch, a project manager for Auger at the Fermi National Accelerator Laboratory in Batavia, Illinois, says: "We're really optimistic that when the time comes to make up what's missing, we can do that."

Nonetheless, Auger's participants continue to worry. At the April meeting of the American Physical Society in Albuquerque,



Cosmic detection: a surface detector (left) and telescope camera (above) in place in Mendoza.

New Mexico, physicist Michael Turner of the University of Chicago urged US funding agencies to "ensure timely completion and operation" of the observatory in case there was a problem. "The whole future of this field hinges upon the result of the southern Auger," he said.

♦ www.auger.org

PIERRE AUGER OBSERVATORY

Mummy's silence broken at last

Rex Dalton, Tucson

The lock on a 2,000-year-old mummy's scientific secrets has been opened by a South African archaeologist — after a three-year wrangle with local tribes and landowners over its custody.

Johan Binneman, archaeology curator at the Albany Museum in Grahamstown, found the mummy in 1999, during an approved expedition at a hunter-gatherer shelter site in the Kouga Mountains in South Africa's Eastern Cape province.

But government officials refused to let Binneman, whose work was publicly funded, present scientific data on the Kouga mummy until last month, when he attended the biennial conference of the Society of Africanist Archaeologists at Tucson, Arizona.

Last September, Maryna Steyn, a physical anthropologist at the University of Pretoria who worked with Binneman, was unable to present a manuscript on the specimen at the fourth World Congress on Mummy Studies in Greenland, pending approval by South African government officials.

The mummy is 145 cm tall and is thought to be a male of 50–55 years of age, possibly a shaman or a tribal chief. It was found in a near-fetal position — with its feet bound — and wrapped in leaves of a poisonous plant (*Boopha disticha*), a well-known coagulant.



Habeas corpus: disputed ownership has delayed scientific description of the Kouga mummy.

Researchers say that this discovery is the first of its age in southern Africa.

As well as problems with landowners, the find set off a tussle between local tribes. Two, the Inqua and the Gonaqua, laid claim to the mummy as a former leader. But provincial government officials — many of them from a third tribe, the Nguni — would not accept their claims.

"The mummy symbolizes our culture for centuries," says Jean Burgess, chief of the Gonaqua, claiming that provincial officials had dragged their feet in allowing its full examination. The officials could not be reached for comment.

Binneman says the disputes are deterring archaeologists from exploring the Kouga Mountains site further.

ALBANY MUSEUM



Worldly blues: protesters urge delegates at the Bali talks to do more to protect the environment.

Plans for a strategy to save the environment run into the sand

Jakarta Talks to hammer out an international action plan for protecting the global environment while encouraging economic development have ended in impasse. The Indonesian island of Bali last week hosted the last preparatory meeting for the World Summit on Sustainable Development, due to open in South Africa on 26 August. But delegates left having reached deadlock on the use of aid money to relieve poverty, and on widening access of products from developing countries to developed-world markets.

The Bali meeting was the fourth attempt to develop a plan before the Johannesburg gathering, which has been billed as the most important environmental meeting since the 'Earth Summit' in Rio de Janeiro in 1992.

A spokeswoman for the UN Division for Sustainable Development remains optimistic that the problems could be ironed out in Johannesburg. "That's the dynamic of these negotiations," she says. "When governments know there is still time to get an agreement they will use that time."

Contaminated feed leads to Israel's first BSE case

Jerusalem Israel reported its first case of bovine spongiform encephalopathy (BSE) last week, becoming only the second country outside Europe — after Japan in 2001 — to report the disease in locally bred cattle.

The 10-year-old Holstein cow had shown BSE symptoms for two days before it was slaughtered. The diagnosis was confirmed by a Swiss reference lab working for the World

Organisation for Animal Health.

The most likely source of the infection is poultry meat-and-bone meal contaminated with infected mammalian protein, imported from Europe in the 1990s. Israel stopped importing mammalian meat-and-bone meal, now known to be the source of the disease in Europe, from Britain in 1988 and from other countries in 1990. The Israeli Ministry of Agriculture says that all slaughtered cows aged over 30 months will now be tested for BSE before being sold for meat.

It has been estimated that some 50 further countries are at risk from the disease because they imported meat-and-bone meal from Europe in the early 1990s.

Framework passed as objection is dropped

Munich Ministers from the member states of the European Union (EU) last week approved the sixth five-year Framework research programme — without the expected protests from countries that are nervous about funding research on human embryonic stem cells.

The 17.5-billion-euro (US\$16.5-billion) programme had been threatened by Germany, where rules on such research are more restrictive than those agreed by the European Commission, allowing research only on cell lines created before 20 January 2002 (see *Nature* **415**, 566; 2002).

Less than 50 million euros, or 3% of the total sixth Framework budget, is for research that could involve human embryonic stem cells. But late last month, Germany rallied the support of Italy and Austria to oppose approval of the entire programme. The three countries would have represented a blocking minority under the EU's voting rules, precipitating lengthy mediation that could have delayed the next round of EU research funding for many months.

But last-minute negotiations led by Spain, which currently holds the EU presidency, persuaded Italy to drop its objections, and cleared the way for the sixth Framework to be launched in November.

Bush asked for \$200 million more for global disease fund

Washington The US Senate has asked President George W. Bush to donate an extra \$200 million to the Global Fund to Fight AIDS, Tuberculosis and Malaria.

The request was added to a \$31.5-billion supplementary spending bill that is intended primarily to support the 'war on terrorism'. As the version approved by the House of Representatives includes the same wording, the extra money is now likely to be added to the \$500 million the US administration has already committed to the fund.

The global fund gave out its first round of

grants, totalling \$378 million, at its inaugural board meeting in April in New York (see *Nature* 416, 773–774; 2002). But its total endowment of some \$2 billion, contributed by various countries, organizations and individuals, has been criticized for being too small to make an impact on the global burden caused by the three diseases.

Marine atlas to aid ocean management

Paris A new web-based atlas aims to become an oracle for policy-makers charged with managing marine resources.

The UN Atlas of the Oceans, launched on 5 June at UNESCO's headquarters in Paris, links to hundreds of maps giving information on fishery resources and ocean characteristics such as salinity, volcanic activity, ice cover and navigation routes. It was created by UN agencies collaborating with the US National Oceanic and Atmospheric Administration and Russia's Head Department of Navigation and Oceanography, which have provided access to maps and databases.

The atlas's editors say that information will be constantly updated for the benefit of policy-makers. But experts think it unlikely that the atlas will remain timely and detailed enough for use as a research resource by marine scientists.

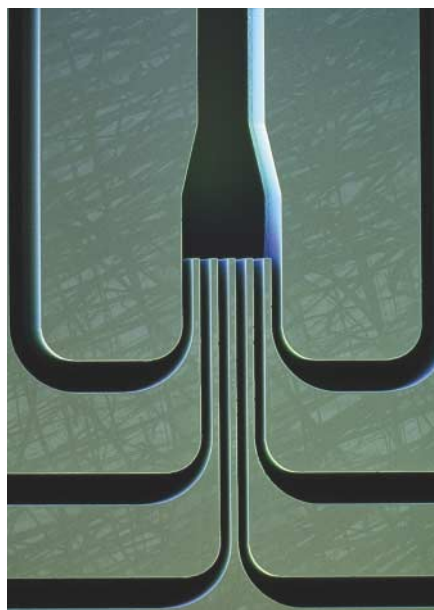
► www.oceansatlas.org

Nanotechnology panel calls for more collaboration

Washington A panel of experts convened by the US National Academy of Sciences has advised the White House to set up a special fund to support interdisciplinary research in nanotechnology. The recommendation comes after the panel examined the National Nanotechnology Initiative, which was launched in 2000 and has roughly doubled federal funding for nanoscale research to some \$500 million per year.

The panel, chaired by Samuel Stupp of Northwestern University in Evanston, Illinois, says that the initiative is going well, but needs to provide greater incentives for genuinely interdisciplinary work. Although the panel's report does not identify any particular missed opportunities, it is enthusiastic about biomedical applications, including nanoscale biosensors and drug-delivery systems, and biocompatible nanomaterials for medical implants.

The report suggests that a special fund managed by the White House Office of Science and Technology Policy (OSTP) is needed to encourage "meaningful interagency collaborations" between the National Institutes of Health, the National Science Foundation and the Department of Energy. This would be an unusual step — the OSTP normally leaves



Small wonder: a DNA-analysis microchip. Will nanoscale devices also find uses in biology?

research funding to the agencies but seeks to coordinate their activities.

Tiny aircraft flies on a shaft of light

Tokyo It sounds like the ultimate in classroom one-upmanship: a laser-powered paper plane. But the device — actually made from folded aluminium foil — may one day lead to tiny inexpensive aircraft for monitoring climate or volcanic eruptions.

Takashi Yabe of the Tokyo Institute of Technology and his colleagues launched their plane by firing a blast of light from a commercial laser at the plane's 'fuel': either an acrylic polymer coating or a droplet of water (T. Yabe *et al. Appl. Phys. Lett.* **80**, 4318–4320; 2002). The aircraft took off into a gracefully curving flight at a speed of around 1.4 metres per second.

Yabe's team is revisiting an idea first proposed 30 years ago. In 1972, Arthur Kantrowitz of Dartmouth College in Hanover, New Hampshire, pointed out that a laser beam focused onto a target could produce a jet of gas that would push it forward, just like a jet engine.

Yabe and his colleagues cannot yet control or power their aircraft in flight. But if the liquid droplet were replaced from a reservoir, sustained flight might be possible. It might be steerable by aiming the laser at fuel on one wing, or the wings could be made of alloys that change shape when heated by the laser.

Correction The News report on the annual meeting of the International Whaling Commission (*Nature* **417**, 476; 2002) erroneously reported that the United States and Russia were requesting a yearly quota of 280 bowhead whales for aboriginal communities. This was in fact the total request for 2003–2007.

Plugging the brain drain

China produces fine scientists, but too many go abroad for training and do not return.

David Cyranoski visits a programme that aims to give scientific high-fliers a reason to stay put.

Liqun Luo is in demand. It's seven in the morning and a queue is already forming for his cell-biology class at the Shanghai Institutes for Biological Sciences (SIBS). Luo, a visiting lecturer, is not due for another hour and a half, but the students have arrived early to get the best seats.

Back in his laboratory at Stanford University in California, Luo's research team jokes about his new rock-star status. But, like real rock fans, many of Luo's Chinese students want to follow in his footsteps. Young Chinese who are serious about science often feel that they have to train abroad — and most do not return. Luo's classes are designed to persuade them to stay.

The US National Science Foundation estimates that around 33,000 Chinese students enrolled in graduate programmes in the United States in 1999. And of Chinese students receiving their doctorates that year, almost 90% said that they planned to stay in the United States. If China is to realize its ambitious plans to expand life-sciences research (see *Nature* **410**, 10–12; 2001), it must halt this brain drain.

Stimulation on a shoestring

Luo's lectures are part of a new graduate programme in the biological sciences that is bringing world-class education to China, and might begin to stem the flow. The budget of US\$20,000 a year is just enough to cover guest lecturers' expenses. But with this shoestring funding the organizers are creating a course that is stimulating students and stirring up the Chinese education system.

It is all the idea of Yi Rao, a neuroscientist at Washington University in St Louis, Missouri, and Jiarui Wu, a molecular biologist at SIBS. Wu gained his PhD at the Swiss



Wake-up call: expatriate scientists lecturing on the latest research have proved a hit with students.

Federal Institute of Technology in Zurich, and then carried out research at the State University of New York in Syracuse. "After getting that great education overseas, I felt a duty to contribute to first-class graduate education here," he says.

Together with Rao, he persuaded the Chinese Academy of Sciences to fund a one-year programme of graduate lectures in molecular and cell biology. The courses, which began in 2000, are taught in Chinese but the outlook is global. All the 25 lecturers have worked outside China. Fifteen are based in the United States and return for a week each year to teach classes.

The course is an eye-opener for students used to lecturers working from textbooks and papers in Chinese journals. "Here the professors show us their latest research, so we know it is cutting-edge," says Ning Li, a first-year student. The students also appreciate the relaxed atmosphere. "In a traditional class we cannot ask the professors many questions," says second-year student Yan Li. "We do not want to offend them."

Lecturers on the programme also make students examine scientific papers with a critical eye. Rao, for example, challenges his

students to criticize a *Nature* paper. "The students used to worship papers in journals like *Nature*, *Science* and *Cell*," says Rao. "It is important that they think about the content of a paper and not get carried away by its publication in a fashionable journal."

Will it work?

Like any new course, the programme has glitches. Some students complain that graduates of previous courses who run discussion sessions cannot explain all the material they have to deal with. Others say that the content of the different classes is hard to tie together. And lecturers acknowledge that the class size of 200 is too big to allow students easy access to the teacher.

These problems do not seem to be dampening the students' enthusiasm, but is the programme going to persuade them to stay? Despite Nang Li's praise for the course, he is in no doubt about what he wants to do next: "I will go to the United States to do postdoctoral work." Many students say the same.

Comments such as these do not surprise the lecturers. "I don't think a good teaching programme alone can keep more top researchers in China," says Bai Lu, a neuroscientist at the National Institute of Child Health and Human Development in Bethesda, Maryland, and a lecturer on the course.

But the Chinese government is pouring money into biological science and seeking to commercialize the results of research. If the SIBS programme can be emulated elsewhere in China as well, it could provide another reason to stay. And if it generates a flow of talented home-grown postgraduates, it might even convince more expatriates to return to collaborate with this highly trained workforce.

David Cyranoski is *Nature's* Asian-Pacific correspondent.



Liqun Luo (left) and Yi Rao are bringing world-class education to Chinese students.



The great primate debate

When can invasive experiments on monkeys or apes be justified? And what would be the consequences for biomedical research if they were to cease? Sally Goodman and Erika Check pose some difficult questions.

There can be few more emotive issues in science than experimentation on our fellow primates. To the researchers involved, such research is vital for continued scientific progress against killers such as HIV and Alzheimer's disease. To militant animal-rights activists, those who work on primates are the worst kind of animal 'abuser', deserving public vilification.

Public attitudes and regulations vary from country to country. In Britain, for instance, research on primates is highly unpopular, and tightly controlled; whereas in the United States, both public opinion and the regulatory climate are more permissive. This international variation provides an opportunity to ask some key questions about the current and future practice of scientific research on primates. Could some of the experiments that are being conducted be abandoned without significant loss? Has the unfavourable climate in countries such as Britain already impeded scientific progress?

And against this background, how might the debate over research on primates develop in the future?

Given concerns for the animals' welfare, and the fact that many primates are threatened or endangered, researchers everywhere agree that invasive experiments on primates are only justified when there is no alternative. "We should only use primates where we can be sure that our research results will be used by clinicians," adds Bert 't Hart, an immunologist at the Biomedical Primate Research Centre at Rijswijk in the Netherlands.

The difficulty of procuring primates (see 'Supply and demand', below) and the high cost of keeping them also help to ensure that such experiments are not conducted lightly. But the fact that the United States, which used some 57,000 primates in research in 2000, uses more than five times as many of the animals as the European Union suggests that the interpretation of 'no alternative' differs from nation to nation.

Supply and demand

Throughout the world, there is a shortage of primates for research — and nowhere is this more acute than in Europe. More than half of the primates used for research in the European Union come from farms in China, Mauritius, Israel and the Philippines — a situation that satisfies no one. Chinese farms, for instance, are depleting wild populations because they mainly supply first-generation offspring from wild-captured animals. And, after a campaign led by animal-rights groups, no airline will fly primates directly into Britain, Europe's biggest user. This has had little effect on the number of primates imported, but it means that some monkeys now endure a combined air and land transit of more than 50 hours.

Researchers working on primates argue that the solution is to

establish sustainable breeding colonies in Europe. The Dutch government in April agreed to provide 40 million euros (US\$37.5 million) to the Biomedical Primate Research Centre in Rijswijk to improve its research facilities and animal housing, to look into alternatives to primates in research, and to expand its breeding programme. But that will not solve the supply problems facing researchers elsewhere.

In Japan, where the Japanese macaque (*Macaca fuscata fuscata*) is widely used in neuroscience research, some labs rely in part on wild monkeys caught by local authorities that regard them as agricultural pests. But pressure from animal-rights activists to stop the supply to researchers means that the monkeys are now often killed instead.

The situation in the United States

— where there is federal support for macaque breeding programmes, and there are several private-sector suppliers — is somewhat better. But demand still outstrips supply, particularly for the pathogen-free rhesus macaques needed for AIDS research, and researchers complain that government support for primate breeding fluctuates from year to year. Demand is also expected to rise, as funding pours into research to combat bioterrorism, some of which will involve primate models.

Perhaps the biggest worry is that many primates may disappear from the wild as a result of habitat destruction and hunting. Greatly expanded captive breeding might be the animals' best long-term hope of survival. That could, of course, also provide a stable supply of primates for research.



But don't bet on the scientists involved and their opponents putting aside their differences to collaborate on conservation through captive breeding. "The danger with supporting breeding centres for research is that, if animals are available, then scientists will find uses for them," says Maggy Jennings, head of the research animals department at Britain's Royal Society for the Prevention of Cruelty to Animals.



Relative values: experiments on primates have sparked many protests in Britain.

Britain, generally seen as having the world's toughest regulatory framework, is a good place to start examining the issues. And the British situation should be brought into clearer focus this week by a new series of papers from the Boyd Group, a discussion forum that aims to find common ground between biomedical scientists and antivivisectionists. It was formed in 1992 by neuroscientist Colin Blakemore of the University of Oxford and Les Ward, director of Advocates for Animals, a pressure group based in Edinburgh.

Published by the British Psychological Society, the papers cover such questions as the 'moral status' of monkeys and apes — essentially, whether their capacity for social and mental suffering should single them out for special protection, or even certain 'human' rights. They also include a discussion of the use of primates in toxicity testing for drugs, which accounted for three-quarters of the primates used in licensed experiments in British labs in 2000.

Most regulatory bodies require drugs to be tested on one rodent and one non-rodent species before entering human clinical trials. Usually, dogs are used, but toxicologists may also turn to primates. In the case of protein- and peptide-based drugs and vaccines — which are likely to grow in importance in the postgenomic era — there may be no alternative. In such cases, a difference of a single amino acid can mean the difference between saving a cancer patient and causing liver damage, so toxicologists want a model that is as close to people as possible.

For other types of drug, there is growing evidence that models such as mini-pigs — dwarf varieties that can be kept conveniently in the lab — might be just as good as monkeys.

But according to the Boyd Group, some companies are still using primates when dogs do not fit the bill because they fear that regulators, who have little experience with alternative models, might send back the application with a request for monkey data.

Better by design

The Boyd Group advocates more discussion and negotiation with regulatory authorities early in the design of toxicity studies of new drugs to ensure that primate use is really justified. Other experts argue that there is potential for reducing the number of primates used by giving more consideration to experimental design and statistical methods. But the Boyd Group's papers caution that a sudden move to ban primate use in Europe might simply lead to the work being relocated to countries where regulations are less restrictive, such as China.

Toxicology is not the only area where some researchers feel that there is scope to reduce primate use. Pierre Druilhe, a malaria researcher at the Pasteur Institute in Paris, believes researchers in his field sometimes opt for primate models for fallacious



Colin Blakemore (left) and Les Ward are seeking common ground for science and animal rights.

anthropomorphic reasons. The relationship between host and parasite is the product of countless generations of co-evolution, Druilhe notes. So for certain aspects of the work, it may be better to study a naturally occurring host-parasite interaction in a rodent than, for example, infecting monkeys with the human malaria parasite.

But similar evolutionary arguments also mean that primate models are the only viable option for research on some infectious diseases. Research into HIV vaccines is a case in point. Here, pathogen-free rhesus macaques (*Macaca mulatta*) are regarded as an essential model. When infected with simian immunodeficiency virus (SIV), a close relative of HIV, they become ill very rapidly. The symptoms closely mirror human AIDS.

"Monkey models are particularly valuable when you want to introduce genetically defined viruses into the animal and ask what are the consequences," says Joe Sodroski, director of the Center for AIDS Research at the Dana-Farber Cancer Institute in Boston. And in recent years, researchers have developed genetically engineered versions of SIV dubbed SHIV, containing genes for proteins from HIV that are recognized by the immune system. This has made the rhesus macaque even more suitable for vaccine studies.

One important insight from monkey research has been that some vaccine strategies that are applied to other diseases — those that stimulate primarily an antibody response — seem unlikely to work against HIV. One such vaccine has entered clinical trials, and few AIDS researchers rate its chances of success. But without macaque models, researchers might still be making intense efforts in this direction, potentially wasting large sums of money and exposing human volunteers to the risks of side effects while failing to protect them from infection. Significantly, several vaccine strategies that were developed in macaques are now moving into human



Without primate research, would candidate vaccines for HIV now be ready for trials?



Viral insight: rhesus macaques offer an important model for studies of HIV.

studies using non-invasive techniques such as functional magnetic resonance imaging, which can pinpoint regions of brain activity, electrophysiological recordings from monkeys' brains have transformed our understanding of cognitive neuroscience. Primate research is also refining neurologists' ability to interpret human brain images, and might bring important insights into a range of human diseases. "Virtually every major neuropsychiatric disorder and many neurosurgical and neurological ones involve the frontal lobes," says a British neuroscientist. "These are areas that are well developed only in primate brains."

Public unease

But despite the potential medical applications, many fundamental neuroscience studies seem arcane to non-specialists. And with images showing primates with electrodes protruding from their skulls being among the most effective recruiting tools for the animal-rights movement, neuroscientists who work on primates have an extremely difficult job in trying to win public acceptance for their work. "I suspect it is a debate that cannot be won, as it cannot be argued in rational terms only," says one British researcher.

The British public is known for its unease about animal experimentation, and the country is also home to a vicious minority of animal-rights terrorists, who have targeted scientists with letter bombs, fire bombs and envelopes booby-trapped with razor blades.

clinical trials — which means that researchers will soon be able to assess in detail the accuracy of the macaque model by comparing the results of the monkey and human trials.

Primate research is also seen as unavoidable in other areas of research with obvious medical applications. Take reproductive biology: research on new forms of oral contraception, for example, can be undertaken with rodents or sheep, because the pituitary gland at the base of the brain, responsible for the release of reproductive hormones, is broadly similar in all mammals. But attempts to understand the underlying genetic basis of human infertility rely heavily on primates.

Macaques are also vital for research into endometriosis, a painful and distressing abnormality of the menstrual cycle, because lower mammals do not menstruate.

For neuroscientists, the very attributes that make primates such an essential model — their high cognitive ability and similar brain architecture to humans — is precisely why invasive experiments are so controversial. Recent studies in macaques based on electrical recordings from individual neurons, for instance, have been key to advancing understanding of how the brain makes simple decisions based on its perception of the outside world. In combination with

Too close for comfort?

Few people can look at a chimpanzee in the eye and fail to feel a sense of kinship. And in the European Union, invasive research on the species will soon be a thing of the past. In Britain, no chimp experiments have

been conducted for many years, and the practice was formally banned in 1997. The final curtain is now falling, as the Dutch government has decided that research on chimps at the Biomedical Primate Research Centre (BPRC) in Rijswijk, the only lab in the European Union still using the animals, will stop by December.

Japanese academics have already halted invasive chimp research, and are pressing for a total ban — the country's drug-industry labs currently hold 107 chimps, used for trials on hepatitis C and malaria. But it is in the United States that chimps have been used most extensively. In the early days of AIDS research, scientists found that chimps can be infected with HIV, and hundreds of the animals were pressed into service in studies of the virus. But it soon emerged that studying a related virus, called SIV, in macaques was far more useful.

"Chimps are an endangered species. They are extremely costly to do experiments on. And HIV in general does not cause AIDS in the chimpanzee," explains Norm Letvin of Harvard Medical School in Boston, a pioneer of the macaque/SIV model.

In the late 1990s, interest in chimp models of AIDS temporarily resurged when Frank Novembre of Emory University in Atlanta identified an HIV-infected chimp that did develop AIDS. Controversially, he infected several other chimps with the same virus. But in 1999, the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, convened a meeting to consider the issue, and recommended that US researchers should not infect any more chimps with the virus. "We had an interesting finding, but I understand the reasoning behind not pursuing it," says Novembre. The other

chimps he infected show signs of immune-system damage but have not yet died of AIDS.

Today, chimp research in the United States has largely stopped, although some experiments continue — mostly on hepatitis C, as no other animal can be infected with the virus. Jon Heeney, a virologist at the BPRC, argues that this is one area where halting research entirely would be damaging. Chimps infected with hepatitis C do not become ill, and Heeney believes that researchers are close to understanding why — which might lead to treatments for humans.

But even if all experiments end, the legacy of US chimp research will live on in the form of 1,500 animals, many with HIV, that are surplus to requirements. Their fate has become a *cause célèbre* for the animal-rights movement, and only now are government agencies beginning to think about their future care.





British neuroscientists are working on marmosets (above) but would like to switch to macaques.

Blakemore, who has worked on visual development in rodents, cats and monkeys, has been forced to live under a police guard for many years.

In the United States, advocacy by patients' groups has so far outweighed the impact on the public consciousness of the animal-rights movement. "Our government and the people in this country are very supportive of biomedical research and you don't have to think about it too long before you appreciate the importance of doing some of those research studies in animal models," says Mark Feinberg, a vaccine researcher at Emory University in Atlanta who works on macaque models of AIDS at the Yerkes National Primate Research Center.

But in parts of mainland Europe, where the British experience seems less remote, scientists worry about the future climate of public opinion. Germany, for instance, also hosts an active animal-rights lobby. Nikos Logothetis of the Max Planck Institute for Biological Cybernetics in Tübingen is one of the world's leading neuroscientists, and says his primate research has not so far been hampered by excessive regulations or public opposition. But he does worry about "irrational restrictions in the future" that might force him to move abroad.

The difference in the number of primates used for research in Europe and the United States already suggests a major transatlantic schism. But most researchers contacted by *Nature* believe that, with a few exceptions — research on chimpanzees (*Pan troglodytes*) being the obvious example (see 'Too close for comfort?', opposite) — there are no major qualitative differences between the types of primate experiments that are sanctioned in the United States and in Europe. Rather, the difference mainly reflects the availability of funding and animals, the greater research activity of the US drugs industry, and the differing regulatory hurdles that must be overcome before a project is given approval.

US researchers often complain about the red tape associated with animal experimentation in general. But for rhesus macaques, the requirements are similar, in essence, to

those for experiments on other laboratory animals. In many European countries, however, researchers must go to much greater lengths to demonstrate the special justification for using a primate model.

Scrutiny is particularly stringent in Britain, as one researcher investigating surgical approaches to treating Parkinson's disease can attest. Since 1997, he has been working on an experimental model of this neurodegenerative condition in macaques, which develop Parkinsonian symptoms after being given a neurotoxic drug called MPTP. But six months ago the government halted his work, demanding further justification for the research and imposing modifications to experimental procedure.

Help or hindrance?

Some researchers argue that Britain's tough line brings some advantages. Christian Keyzers, a neuroscientist at the University of Parma in Italy, who has also worked in Britain and the United States, believes that strict legal requirements have made British technicians more proactive in addressing animal welfare.

But according to others, a tough regulatory stance, combined with public antipathy and outdated facilities, is placing countries such as Britain at a disadvantage. The University of Cambridge, for instance, wants to build a new £24-million (US\$35-million) centre for neuroscience, which would conduct research on macaques. But in February, the local authority denied planning permission, citing the likelihood of public-order disturbances caused by animal-rights protesters. In the meantime, Cambridge scientists are working with common marmosets (*Callithrix jacchus*) — which are not such a good model because of their smaller brains. One neuroscientist who recently moved to

Britain believes that, as a result of such difficulties, the country is perhaps a decade behind the United States in studies of the function of neural circuits in living animals.

Such concerns help to explain why UK Prime Minister Tony Blair last month made a speech urging the British public to support developments in science, highlighting the stalled Cambridge primate facility (see www.nature.com/nature/blair.html). But with the British Union for the Abolition of Vivisection releasing a video the next day alleging that marmosets that had undergone brain surgery at Cambridge suffered bleeding head wounds, fits, vomiting and whole-body tremors, it is clear that British primate researchers face an uphill battle in winning hearts and minds.

So how is the debate likely to proceed? Arguments about whether some of the primate experiments that currently take place could be abandoned will no doubt continue. But if we are not to sacrifice medical developments that, for now, can only be achieved using primates, some research must go on. And unless there are any major shifts in attitudes, this work seems increasingly likely to migrate from countries such as Britain to the United States and other nations where the climate of public opinion is more favourable.

The real crunch will come if the public attitudes and violent protests that prevail in some European countries cross the Atlantic. "There is already a movement in this country to stop research on non-human primates," says John VandeBerg, director of the Southwest National Primate Research Center in San Antonio, Texas. "We feel very threatened that the events that have taken place in Europe could take place in the United States."

Sally Goodman is *Nature's* French correspondent; Erika Check is *Nature's* Washington biomedical correspondent.

♦ www.boyd-group.demon.co.uk



US researchers fear that protests, such as this in Washington, will grow and restrict their work.

Data security is crucial for Japanese science

Rules that are based solely on voluntary guidelines will not gain the trust of the public.

Sir—The Japanese parliament is currently discussing new data-protection legislation, the first to be introduced in Japan, which will provide scientists with a broad exemption from the proposed rules. This exemption has not been widely discussed, and few citizens are even aware of the issue. Although the biomedical research community and pharmaceutical industry initially indicated tacit consent, concerns are now being voiced.

We believe the bill does not constitute an adequate regulatory framework for privacy protection in biomedical research. Several ministries, agencies and scientific societies have, in the past, drafted guidelines that have increased awareness of ethical issues, such as informed-consent procedures and review by research ethics committees. Yet the scope of the existing guidelines is restricted and there is considerable confusion about how the rules would be applied. We fear that data protection based solely on voluntary guidelines is insufficient to win public trust.

Progress in large-scale human genome research has dramatically increased the amount and content of personal data used in many areas of biomedical research. Large online databases linking personal

information with genomic and clinical data are now feasible. For example, BioBank UK (see *Nature* **417**, 9; 2002) is seen by many observers as an important step forward for clinical-genomics research in Britain. But its success will depend on the participation of UK citizens. Such support is unlikely without a well-conceived scientific strategy and, equally important, a convincing legal and regulatory framework for privacy protection.

Assembling large-scale cohorts for genome research remains very difficult in Japan; we believe a more widespread sharing of resources is now inevitable. Yet without a coherent, transparent approach to data security and data protection, the Japanese public is unlikely to support efforts to build large databases containing medical information and lifestyle details linked with genetic samples, because of distrust and the perception that Japan's medical and scientific community is unable to govern itself.

An inadequate data-protection regime could harm progress in clinical genetics and other biomedical sciences in Japan for many years. The result could well be that Japanese scientists and enterprises go

elsewhere — already, Japanese companies are choosing Singapore and even Mongolia for new clinical-genomics research facilities. The Japan Association of Bioindustries Executives, a group representing the chief executives of Japan's major biotechnology companies, has just released a statement (www.jba.or.jp/jabex) urging the government to strengthen genetic privacy protection.

We believe that large databases containing both genetic and medical information are inevitable for the future of biomedical research. But without adequate data protection or public trust, such databases are unlikely to be set up in Japan. For biomedical research to progress in Japan we, as members of its scientific and medical community, believe it is time for Japanese scientists to engage in a constructive public debate on this issue.

Eitaka Tsuboi

President, Japan Medical Association, World Medical Association, 28-16, Honkomagome 2-chome, Bunkyo-ku, Tokyo 113-8621, Japan

Other signatories to this letter:

Norie Kawahara (*freelance journalist*)

Tadahiro Mitsuishi (*attorney-at-law*)

Akira Oshima *President, Japan Association of Cancer Registries*

Shohei Yonemoto *Director, Center of Life Science and Society*

Collaboration can work if inequality is recognized

Sir—Your News report “Science collaboration stymied by relentless Middle East conflict” (*Nature* **417**, 209–210; 2002) mentioned my support for Palestinian–Israeli cooperation. I would like to expand on my views. The Palestinian–Israeli conflict is neither a scientific issue nor a personal conflict but a struggle between two peoples over the same land. Therefore neither scientific collaboration in itself nor the personal relationships developed therein can advance the cause of peace unless conflict-related issues are also addressed and resolved.

Israelis have a strong army, resources, freedom and control over the land; Palestinians have none of these. One is occupier and the other is occupied. This inequality underlies any interaction between Palestinians and Israelis, including scientific cooperation, and should be the framework for any constructive dialogue. Unfortunately, this imbalance is often replicated in the collaboration and becomes a cause of frustration and disengagement on the part of the Palestinian partner.

It requires courage and commitment to

widen the partnership beyond the realm of science, to rise above the current polarization and high emotional pitch in our respective societies, to place universal humanitarian values above nationalism and to take a clear stand for justice and peace. Yet it is possible, as demonstrated by individual academics as well as organizations such as *Médecins sans Frontières*, *Physicians for Human Rights* and the *Alliance of Middle Eastern Scientists and Physicians*. An example of such a stand is a recent statement by 300 Israeli faculty members, which can be seen at www.seruv.org.il/UniversitySupportEng.asp.

The international scientific community should become more actively engaged. Foreign collaborators and funding agencies can request a commitment to basic human rights and equality, and to the principle of equal academic freedom and access to education for Palestinians and Israelis. Israeli scientists and institutions should express solidarity with Palestinian universities under siege and use their political clout to assist them. All partners should insist on a return to negotiations based on UN resolutions and international law.

A boycott of Israeli science would be relatively easy but, in the long run, counter-productive. Much more challenging and

important is to use scientific interactions creatively, to promote true reconciliation.

Yehuda Tzfati

Alliance of Middle Eastern Scientists and Physicians and Department of Genetics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Scientific links support an unjust peace process

Sir—Cooperation between Israeli and Palestinian academics since the 1993 Oslo peace agreement (see *Nature* **417**, 209–210; 2002) has been primarily donor-led — that is, imposed or strongly encouraged via the carrot of money in a money-starved environment, with the underlying assumption that such cooperation would assist the peace process.

This push for “Israeli–Palestinian scientific cooperation”, as if science were divorced from society, was regarded by many Palestinian academics as suspect. Most Palestinians opposed a basically unjust and non-sustainable peace process, as recent events in the area clearly confirm.

Palestinian academics have been paying the very heavy price of occupation for many years, both personally and in the

under-development of their institutions. More recently, the conditions of closure and siege have reduced our scientific endeavours to near-paralysis, leaving us unable to teach, let alone conduct research. I believe it would be more rational to work towards preserving Palestinian academic institutions against the Israeli army's onslaughts, and rebuilding them, before collaborating with Israeli academics.

To me, this seems the right moment to act, including endorsing boycotts, instead of turning to the easier yet ineffective paths of building personal relationships, scientific or otherwise. Boycotts have been effective in raising issues and influencing change.

Israeli and other academics need to stand up for the right of Palestinians to scientific and educational development, academic freedom and freedom of speech for all, not simply on one side of the border. Cooperation with individuals may make academics feel better, but is not helpful, and can possibly be harmful, in general terms. Saying no to academic boycotts may mean that academics are not willing to pay a price for their stated ideals. To many of us here, this translates as a *de facto* endorsement of the Israeli government's attempt to destroy 'anything Palestinian', including the academic institutions that these relationships are supposed to assist.

Rita Giacaman

Institute of Community and Public Health, Birzeit University, Palestine

Did an academic boycott help to end apartheid?

Sir — The assertion made by Steven and Hilary Rose in Correspondence (*Nature* **417**, 221, 2002) that the boycott of South Africa by the world's academic communities "was instrumental in ending apartheid in South Africa" is a deception. Apartheid was actually terminated by two pivotal and interrelated political events.

First, the United States Congress, on 29 September 1986, overrode President Reagan's veto and imposed strict economic sanctions on South Africa. Second, F. W. de Klerk was elected president of South Africa on 14 September 1989. Two months later (16 November 1989), de Klerk announced the scrapping of the Separate Amenities Act, then, on 11 February 1990, freed Nelson Mandela from prison. The rest is historical detail.

So if the Roses, and the signatories of their petition, wish to bring Israel even further to its knees, they may need to persuade Europe and the United States to increase by an order of magnitude the stringency of the Arab-led international trade boycott of Israel that has been in place since 1948.

As it stands, the petitioners have not made a cogent argument for why they selected Israel alone — from the many

imperfect nations of the world — for their proposed academic boycott.

George Fink

C/o Scotbrain, 78/22 Levi Eshkol, Tel Aviv 69361, Israel

Violence versus freedom

Sir — I read with great interest your News story about academic cooperation during the recent violence here (*Nature* **417**, 209–210; 2002). Unfortunately, scientific work has become another victim of the violence. A related point is that of freedom of movement.

For many years, Palestinians could travel fairly freely in Israel. Thousands worked here on a daily basis. The increase in terrorist attacks, especially the suicide bombings, has made this freedom of movement unwise and dangerous. Many innocent people, including scientists, suffer as a result. Israelis are forbidden by the Israeli government to visit the areas controlled by the Palestinian Authority, for their own safety. Israelis who visit Palestinian cities are likely to pay with their lives: hardly "freedom of movement".

I hope that we will manage to achieve a more rational, calm atmosphere in this part of the world — but I fear that it will take some time.

Joel Bigman

ELS Photonics, PO Box 252, Nesher 36602, Israel

In support of scientific exchange

The International Human Rights Network of Academies and Scholarly Societies (<http://www4.nas.edu/oia/oiahome.nsf/web/network>) was created to address grave issues of science and human rights throughout the world. It aims to put into practice the professional duty of scientists and scholars to assist those colleagues whose human rights have been — or are threatened to be — infringed, and to promote and protect the independence of academies and scholarly societies worldwide. The basis of the network's activities is the universal declaration of human rights.

The network seeks to promote the free exchange of ideas and opinions among scientists and scholars in all countries and, thereby, to stimulate the development of collaborative educational, research and human-rights endeavours within academies and the institutions with which they are affiliated.

Moratoria on scientific exchanges based on nationality, race, sex, language, religion, opinion and similar factors thwart the network's goals. They would

deny our colleagues their rights to freedom of opinion and expression; interfere with their ability to exercise their bona fide academic freedoms; inhibit the free circulation of scientists and scientific ideas; and impose unjust punishment. They would also be an impediment to the instrumental role played by scientists and scholars in the promotion of peace and human rights.

This statement, although that of a general principle with universal applicability, was prompted by a petition (see *Nature* **417**, 1 and 221–222; 2002) that advocates a moratorium on all grants and contracts to Israel from European cultural and research institutions. The moratorium being advocated, although surely well-intentioned, is misguided and inevitably counterproductive.

We all look forward to an equitable solution to the crisis in the Middle East, with lasting peace and stability for both Israel and the Palestinian Authority. But the strongest impact of a moratorium would, rather than influencing Israeli policy-makers, seriously and unfairly

harm our scientific colleagues in Israel — many of whom have actively promoted peace through collegial engagement and open communication among academic centres in the region.

Arjuna Aluwihare *University of Peradeniya, Sri Lanka*
Claude Cohen-Tannoudji *Collège de France, Paris, France*
Ayşe Erzan *Istanbul Technical University, Turkey*
François Jacob *Institut Pasteur, Paris, France*
John Polanyi *University of Toronto, Canada*
Pieter van Dijk *Former Professor of Public International Law, Utrecht University, The Netherlands*
Edoardo Vesentini *Politecnico di Torino, Italy*
Torsten Wiesel *The Rockefeller University, New York, USA*

Contact: Carol Corillon
 Executive Director, The International Human Rights Network of Academies and Scholarly Societies, 500 Fifth Street NW, Suite 5-530, Washington DC 20001, USA

This statement was issued on 28 April 2002 by the members of the executive committee of the International Human Rights Network of Academies and Scholarly Societies. It was sent privately to some 60 national academies affiliated with the network. In response to a request by *Nature*, the committee (whose members' institutions are listed for identification purposes only) has agreed to its publication — Editor, Correspondence.

My synapses, myself

Do our synaptic connections make us who we are?

Synaptic Self: How Our Brains Become Who We Are

by Joseph LeDoux

Pan Macmillan: 2002. 400 pp. £20 (UK).

Viking: 2002. \$25.95 (US)

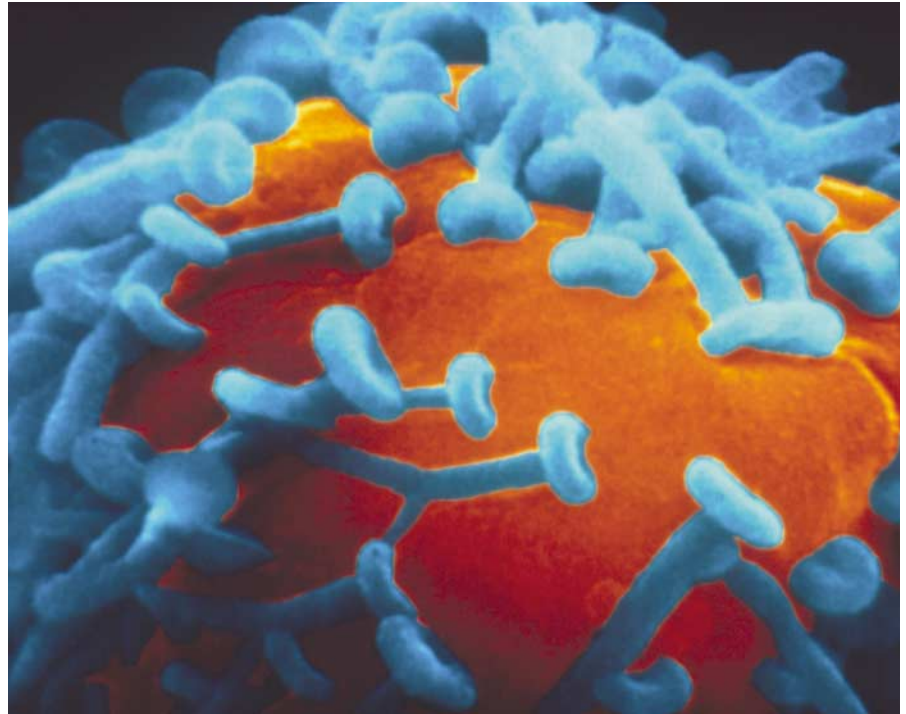
William Calvin

The self, Joseph LeDoux tells us, is “the totality of the living organism”. Most disciplines in the natural sciences focus on only one or two levels of organization. Indeed, Dmitri Mendeleev figured out the periodic table of the elements without knowing any of the underlying quantum mechanics or stereochemistry. There are, however, at least a dozen levels of organization within the neurosciences — and, if we use a metaphor, we temporarily create yet another. This leads to considerable confusion and arguments at cross purposes over whether learning is an alteration at the level of gene expression, ion channels, synapses, neurons or circuits.

Each neuron has thousands of synapses, which produce currents that summate to form an impulse train. But only rarely is the activity of a single neuron sufficient to cause a perception or trigger an action. Neurons usually act as members of ‘committees’ — what Donald Hebb in 1949 called cell-assemblies. Just as in academia, one individual may function in different committees on different occasions. A concept, including any explicit memory that we can talk about, is probably formed by such a committee. Implicit memories (the ones you can’t talk about) are less differentiated — they are part of the ‘feltwork’, together with motivations and emotions, that biases the choice of one’s next act.

In this well-written 400-page appreciation of behavioural neuroscience, LeDoux argues that synapses are the seat of self. He says, in effect, that you are your memories; that it is the uniqueness of an array of synaptic strengths that distinguishes one twin from another. Fair enough, but why not instead focus on one’s unique array of ion channels? Or neurons, because a neuron is the closest thing we have to a computational unit (synapses have to reach a threshold before they have any influence at all)? Or one’s unique arrangement of those overlapping, redundant hebbian committees?

None of these make for a catchy book title, but relating other things to the synapses proves to be a good way of covering a lot of fascinating material at the overlying levels, including a few updates to LeDoux’s earlier book *The Emotional Brain* (Simon & Schuster, 1996). “Our hopes, fears, and desires influence how we think, perceive, and



Committee members? Axons (blue) from many neurons form synapses to one cell body.

remember. A science of mind needs to account for and understand these complex processes,” he writes.

The chapter “Building the brain”, which includes a knowledgeable critique of the selectionist aspects, is about how brains are wired up diffusely and then tuned and pruned by experience. LeDoux addresses memories (and false memories) as studied by psychologists, the amnesias studied by neuropsychologists, the ‘nomadic memory’ theories for how memories ‘move’ from the short-term hippocampus to the long-term neocortex, and then the increasingly understood neurobiological mechanisms for modifying synaptic strength. He devotes three chapters to the mental trilogy of thinking, emotion and motivation, centred on working memory (not really a memory so much as a limited workspace) and executive functions.

The chapter “Synaptic sickness” is more about slow neuromodulators and second messengers than quick synapses. It is a nice history of biological psychiatry and how it now complements psychotherapy. LeDoux points out that there is “an imperfect set of connections between cognitive and emotional systems in the current stage of evolution of the human brain. This state of affairs is part of the price we pay for having newly evolved cognitive capacities that are not yet fully integrated into our brains.” For

those who appreciated the skill with which Antonio Damasio pulled things together in his final chapter on consciousness in *The Feeling of What Happens* (Heinemann, 2000), LeDoux’s final chapter might disappoint; although it covers more ground, it has the feel of a convergence of black boxes rather than the neurophilosopher’s profundity.

At whom is this book aimed? My guess is psychologists — but by covering much of behavioural neuroscience in respectable detail (no one goes unnamed), LeDoux has shifted the focus. Certainly neuroscientists of all sorts will find the book quite useful; those who arrive from molecular and genetic backgrounds are particularly in need of such a book. It could easily be used as a textbook, and is far more readable than most. But the long sections on who did what, when and where are suited only to those breaking into behavioural neuroscience or to historians of science, not to a more general scientific audience. The non-professional reader of psychology, in search of a coherent view of the self, will find it harder going.

Someone, alas, made an obvious effort to ‘jazz up’ the book to appeal to more general readers. But it is only cosmetic: “Golgi and the gap” may be a useful title for the discussion of how reticulated webs were reconceptualized as distinct units separated by synaptic clefts, but “Chemicals are oozing and sparks flying constantly” will mislead the reader who

doesn't realize that the biggest charge in the brain is a mere tenth of a volt — not the stuff that sparks are made of. ■

William H. Calvin is a neurobiologist at the University of Washington, Seattle, Washington 98195-1800, USA. He is the author of *A Brain For All Seasons*.

A complete history of astronomy

Storia dell'astronomia: dalle origini al duemila e oltre

by Giacomo Leopardi & Margherita Hack
Edizioni dell'Altana: 2002. 646 pp. 37 euros.
In Italian

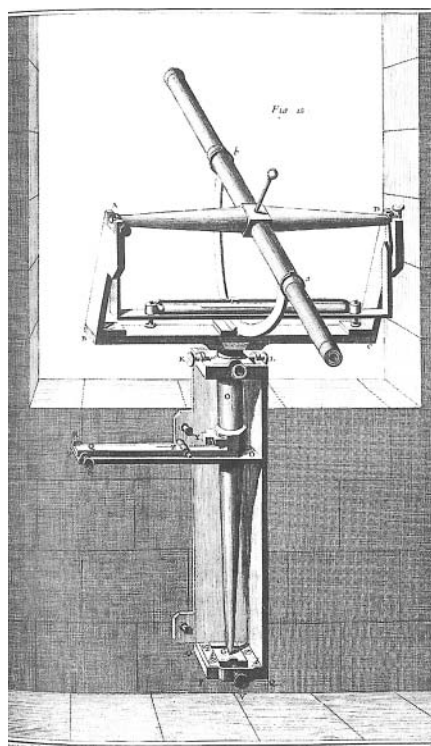
Giovanni F. Bignami

In the introduction to his *Why Read the Classics*, Italo Calvino mentions only one Italian author, Giacomo Leopardi (1798–1837). Calvino does not praise Leopardi as the author of the *Canti*, which is one of the greatest poetic works of the nineteenth century, but instead refers to the unique education that the young genius received in his father's castle. Count Monaldo had a well-organized, 16,000-volume library in Recanati, a small town in central Italy, where young Giacomo used to sit and study in its central room, writing at a small table with his back to the east.

In 1811, while aged just 14 to 15, Leopardi wrote a 300-page history of astronomy, the *Storia dell'astronomia*. It contains more than 1,700 footnotes and references, and an organized bibliography of 300 works spanning two millennia and in many languages, both modern and ancient. He presumably had access to these in the astronomy corner of Monaldo's library. He refers to almost 2,000 astronomers, philosophers, poets and other authors, frequently in the original language. The new edition of his book, updated to the present day by the Italian astronomer Margherita Hack, includes translations of the Greek passages; young Leopardi assumed that his readers would not need them.

However, even at such a tender age, Leopardi was not merely trying to show off his immense erudition. His deep poetic and philosophical mind surfaces frequently in a sea of arid facts, dates and quotations. Consider, for example, his comment on the Latin translation of the Greek poem *Phenomena* by the little-known astronomer Aratus (circa 272 BC). The translation, they say, is by Cicero (yes, the illustrious one), and Leopardi goes off on a tangent on the poetic qualities of the great Roman. Along the way he entirely forgets the original subject — perhaps it was too boring.

Or take the story of the 'mechanical astronomical clock', sent in around AD 800 by the Arabic scientist al-Mansur to Charlemagne,



Big future: improvements to telescopes since the eighteenth century led to astronomical advances.

and Gian Domenico Cassini's observations of Jupiter, showing the giant planet's fast rotation. Leopardi interprets these as implying that man learns to understand what is outside and far away better than what he is sitting on (namely, the copernican debate on the rotation of Earth) or what goes on inside his soul.

In the new, elegant edition of the book, Hack takes over where Leopardi ends. She had a daunting task because so much more has happened in astronomy in the past two centuries than in the previous two millennia. She does a splendid job of presenting an accurate and balanced account, even of today's accumulating, sometimes contradictory and always diverse discoveries. To keep the size of the book manageable, she had to keep within 200 pages, compared with the 300 of Leopardi. She spares us any debate on Cicero's poetics, but has an obvious soft spot for Sir Arthur Eddington, who added so much to our knowledge of the structure of stars, from whom we get ample quotes. However, in a philosophical vein akin to that of Leopardi, Hack ends her contribution with an enlightening reflection on the weak and strong anthropic principles. A difficult but important subject, this is where today's astronomy and philosophy converge to provide a coherent understanding of today's cosmology as it emerges from both ground- and space-based observations.

Hack was one of only a few women astronomers when she began her career, and she has campaigned hard for the place of women in science. The revered elder of

Italian astronomy, she is now an important public and political figure, and celebrates her 80th birthday this week. Happy birthday, Margherita. ■

Giovanni F. Bignami is at the Agenzia Spaziale Italiana, 23 Via di Villa Grazioli, 00198 Roma, and the University of Pavia, Italy.

Portraits of our family tree

The Human Fossil Record. Volume 1: Terminology and Craniodental Morphology of Genus Homo (Europe)

by Jeffrey Schwartz & Ian Tattersall
Wiley: 2002. 400 pp. \$125, £92.95, 147.10 euros

Jacopo Moggi-Cecchi

Interpretations of the human evolutionary process often differ among specialists. The contrasting opinions are in some cases based on a discordant reading of the fossil evidence — the morphological features of the bones and the teeth of our ancestors. Anatomical details described as 'thick', 'broad' or 'deep' by one scholar are not necessarily perceived as such by a second or third scholar and defined using a corresponding term. And this, in turn, may affect the final interpretation.

This book, the first volume in a series of four, was conceived with this issue in mind and is based on an apparently simple idea: to provide a detailed description of the most relevant fossil specimens attributed to the genus *Homo* from all over the world, following a consistent protocol and accompanied by basic photographic documentation. Giving shape to this idea made the project a "Herculean task" (in the authors' own words). The scale of the task and the significance of the book lie in the method used: nearly all of the fossil specimens have been described by the authors, using the same procedure throughout. As a result, readers can be assured that the adjectives used to describe morphological features are applied consistently for each description presented. This, in turn, allows easier, direct comparisons between specimens for each anatomical detail. The descriptive protocol applied and the terminology used are accurately explained, region by region, in the first part of the volume, together with drawings showing the anatomical features mentioned in the text.

The descriptions, arranged by site, range from an isolated tooth to the most representative specimens of large collections, and include recent discoveries up to the two skulls from Georgia. Photos of the specimens (not all of them in sharp focus) have been taken mostly by one author. You won't

find a single measurement in the book: the effort has gone into capturing the shape of a bone or a tooth through its verbal dissection. In an era in which every morphological detail of a fossil can be captured using the most sophisticated technologies, and the information stored in digital archives, a book of anatomical descriptions may seem outdated. This is not so. Detailed morphological accounts are powerful analytical tools and rich sources of information, and this book represents an invaluable archive.

We are at a stage in human evolutionary studies when we are just starting to grasp the developmental mechanisms underlying specific morphologies. When new discoveries of the developmental biology of bone and teeth are made, we will be able to use this archive to document the evolution of a specific shape in our fossil relatives. This is a true thesaurus, in its literal meaning: a textbook to go back to several times, for different scopes.

Jacopo Moggi-Cecchi is at the *Laboratori di Antropologia, Dipartimento di Biologia Animale e Genetica, Università di Firenze, 12 via del Proconsolo, 50122 Firenze, Italy.*

Winning the numbers game

The Honors Class: Hilbert's Problems and Their Solvers
by Benjamin H. Yandell

A. K. Peters: 2002. 496 pp. £28, \$39, 46 euros

W. Timothy Gowers

One of the best ways for a twentieth-century mathematician to achieve immortality was to solve a problem from the famous list of 23 compiled by David Hilbert and delivered to the International Congress of Mathematicians in 1900. The 'honours class' of Benjamin Yandell's title refers to those who managed this feat; his thoroughly researched and highly entertaining book tells us about their lives and about the problems themselves.

Yandell groups the problems according to their mathematical area, more or less as



Sum people? Several mathematicians rose to the challenges set by David Hilbert (top row, centre).

Marine masterpieces

It was the artist Marc Chagall who said: "Great art picks up where nature ends." Now Judith Connor and Nora Deans look at jellyfish and the art that they have inspired in *Jellies: Living Art* (Monterey Bay Aquarium, \$16.95). The fine specimen shown here is *Crossota alba*, which glows in the dark.



Hilbert did himself, although he reclassifies one or two problems because of the unexpected nature of their solutions. He then introduces each area, describes each problem within it, discusses the solution when there is one (by his reckoning only three remain unsolved), and intersperses the mathematics with plenty of biographical information about those who contributed to it. The result is a vivid history of much of twentieth-century mathematics from mathematical, personal and political perspectives — in Russia and Germany, at least, mathematics and politics were not entirely separate, with devastating results.

Occasionally one has too clear an idea of how the book was put together. Unlike many authors, Yandell does not try to hide behind his text. The book is full of such phrases as "Michael [Artin] remarked to me that..." or "[Paul] Cohen thought... and I agree". In addition, some of the anecdotes, although amusing, are not really relevant. In the middle of a long discussion of A. N. Kolmogorov, we are treated to two stories about Sofya Kovalevskaya because she "throws an interesting light on Kolmogorov's early milieu". The remarkable story of Srinivasa Ramanujan is told because he wrongly claimed to have proved a result closely related to the Riemann hypothesis. And the general introduction to the section on number theory begins with a very short paragraph about G. H. Hardy before suddenly being interrupted by three well-worn tales about Paul Erdős. These are included because Erdős has "a connection to the eighth Hilbert problem", but we are never told what the connection is. (The eighth problem is the Riemann hypothesis again; it is one of the three that have not been solved.)

This is the second recent book on Hilbert's problems — the other, Jeremy Gray's *The*

Hilbert Challenge (Oxford University Press, 2000), was reviewed last year in *Nature* (410, 632–633; 2001). However, despite a necessary overlap between these books, they are very different in character and to some extent they complement each other. Gray's book was centred much more on the problems themselves — why they were chosen, why they were so influential, how even very different-looking problems had similar underlying themes — whereas Yandell takes their importance largely for granted.

Gray included biographical details only when they help to explain the progress of the problems, whereas Yandell is far racier and revels in the human story. For example, you will not discover from Gray that Hilbert often became infatuated with pretty women, of whom Emmy Noether (possibly the greatest female mathematician ever) was not one, or that Pavel Aleksandrov liked to work outside under the burning sun wearing nothing but dark glasses and a Panama hat. Yandell's book is also significantly longer than Gray's, which allows him to discuss some of the mathematics in a more leisurely way. He does this well, although some of his discussions fall between two stools: unnecessary for the expert, but incomprehensible to the general reader.

Yandell turned his back on mathematics to become a writer after a distinguished undergraduate career at Stanford University. This makes him an ideal author for a book of this sort, to which he brings a rare combination of mathematical and literary sophistication. *The Honors Class* will be greatly enjoyed by anybody who wants to know more about mathematics and its practitioners.

W. Timothy Gowers is at the Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK.

Health is wealth

Richard G. A. Feachem and
Carol A. Medlin

Economists have long been in the habit of discussing and analysing 'public goods'. A public good is not something that is good for the public (although it can be), but a precise piece of economic jargon. It has two key properties: non-excludability, meaning that it is not possible to prevent the use or consumption of the good by some category of people, such as those who have not paid for it; and non-rivalry, meaning that, for any level of production, the marginal cost of providing the good to an additional consumer is zero.

In both cases, the role of price in regulating the balance between supply and demand breaks down. Classic examples include law and order, defence, and lighthouses. Left to the free market, public goods will be greatly undersupplied because of their non-excludability and non-rivalry. Governments everywhere must therefore step in to finance, subsidize and/or supply them. This is one area of public responsibility that is not contested, even by the most radical free-marketeers.

In medicine and health, factors that cause widespread environmental damage with negative consequences for health are public 'bads', and their reduction is a public good. Obvious examples include ambient air pollution, global warming and ozone depletion. Ambient air pollution typically occurs in a restricted area within one country, such as Delhi, Lagos or Mexico City. The government therefore intervenes by regulation, taxation or other means to provide the public good (cleaner air) that individual citizens cannot buy for themselves. Where air pollution stretches across national borders, such as in the former East Germany, the Czech Republic and Poland, the affected countries may cooperate to reduce a public bad from which they collectively suffer (and to which they collectively contribute).

But some public bads — such as global warming and ozone depletion — affect large regions or even the entire world. For their corresponding public goods, the properties of non-excludability and non-rivalry apply to countries rather than to individual consumers. Ideally, the global or regional reach of these properties will inspire countries to act collectively to stimulate (or discourage) the supply of global public goods (or bads). If global warming could be slowed by collective action, it would be impossible to prevent a country that had not contributed from benefiting, and the cost of that benefit to the dissenting country would be zero. A global public good is clearly something that a 'government of the world' would finance, subsidize or otherwise encourage, if such an entity existed.

Global public goods have recently become fashionable for particular health objectives. However, the term has been widely misused, as if it means 'anything that is good for the health of the global public' rather than its precise economic definition with clear implications for finance and provision. An example is the control of communicable disease on a worldwide scale. Many of the international community's aims in this field have the characteristics of a global public good, such as disease eradication. To fit the definition, the benefits must be enjoyed by all countries in perpetuity, even if they have contributed little, financially or otherwise, to the campaign.

Some important communicable diseases are, by their nature, mainly local in their transmission and epidemiology. Thus, the control of diarrhoeal diseases may have more of the properties of a national public good than a global one. For other diseases, international spread is significant, so control efforts have the characteristics of global public goods. Examples include the control of malaria, tuberculosis and HIV/AIDS, and the (sadly neglected) control of resistance to antibiotics.

Global public goods

Preventing the international spread of disease is not a matter of charity from rich countries to poor. It is a global, collective enterprise from which all countries benefit.

The concept of global public goods for health suggests two lines of argument that are becoming increasingly commonplace. First, the great majority of official development assistance, commonly called 'aid', from wealthy countries to poorer ones has focused on specific projects that benefit specific groups in specific countries. Only a small fraction of the funds has been spent on investment in global public goods, which arguably have had higher returns and a greater success rate than project-by-project assistance.

The second argument is that global public goods for health, particularly the control of communicable diseases, should be removed from official developmental assistance. Preventing the rise of these diseases is not a matter of charity from rich countries to poor ones. It is a global, collective enterprise from which all countries should benefit. Just as the United States invests in its Centers for Disease Control and Prevention (CDC) in Atlanta for the national good, so the world needs to invest collectively and strenuously in practical controls for infections, in slowing or preventing the development of antibiotic resistance, and in research into new drugs. The international surveillance of 'flu strains and regular production of new vaccines against the disease on a regular basis are good examples of a successful collaboration for a global public good, although it mainly benefits wealthy countries in which 'flu vaccines are most widely used.

If a 'government of the world' existed, it would surely invest in a global version of the CDC, not for reasons of social equity or income transfer (aid), but for the welfare and economic prosperity of all people. As we have no such government, nothing quite like this has happened, although the newly created Global Fund to Fight AIDS, Tuberculosis and Malaria is a big step in the right direction. ■
Richard G. A. Feachem and Carol A. Medlin are at the Institute for Global Health, University of California, San Francisco and Berkeley, 74 New Montgomery Street, San Francisco, California 94105, USA. R.G.A.F. is executive director designate of the Global Fund to Fight AIDS, TB and Malaria.

FURTHER READING

Report of the Commission on Macroeconomics and Health (WHO, Geneva, December 2001).

Working for well-being: AIDS sufferers in Bangkok queue up for free health advice and medicines.

Rocks that go bump in the night

Derek C. Richardson

The planets were probably created by collisions between smaller rocky bodies over many millions of years. The identification of a recently formed asteroid family will tell us much about the dynamics of these collisions.

Between the orbits of Mars and Jupiter lie the remains of the early, violent history of the Solar System. These are asteroids, rocky bodies that range in diameter from a few hundred metres to just under 1,000 km. Over the past century, astronomers have realized that asteroids (and comets, their icy cousins originating further out in the Solar System) are the building blocks of planets. On page 720 of this issue, Nesvorný and colleagues¹ report their discovery of an asteroid family that offers unprecedented insight into the dynamics of asteroid collisions — and hence into how the planets of the Solar System formed.

For planets to grow in size from their humble beginnings, they must be continually bombarded by smaller bodies for tens to hundreds of millions of years. This process still happens today: not only do these small bodies continue to strike the Earth, with effects ranging from spectacular fireballs to mass extinctions, but they continue to bump into each other as well. In the past these bumps were relatively gentle, which allowed the first protoplanets to form. But as giant planets such as Jupiter came into existence, the gravitational fields they generated stirred up the smaller bodies. This energy increase made them more likely, if they collided, to break apart than grow, with the result that only the largest protoplanets survived to form planets while the rest were ground down to small fragments.

Nesvorný and colleagues¹ have uncovered the aftermath of one such energetic collision. What makes their discovery unique and exciting is that a large-scale break-up event has, for the first time, been precisely dated. Piecing together the fragments, Nesvorný *et al.* find that a 25-km asteroid in the Outer Main Belt between Mars and Jupiter broke up in collision 5.8 million years ago — a mere blink of an eye in the 4.5-billion-year lifetime of the Solar System.

Theorists studying planet formation would like nothing better than to smash two asteroids together and see what happens. Many questions could be answered. Under what conditions does an asteroid break apart? What happens to the debris? How do pre-existing fractures affect the outcome? Could a 'doomsday' asteroid threatening Earth be stopped by brute-force methods such as a nuclear blast? Of course, full-scale experiments are not possible, so planetary scientists



Figure 1 Aged asteroid. One of only a handful of asteroids that have been visited by spacecraft, *Ida* (pictured here) is a member of the Koronis family that formed over a billion years ago as a result of a catastrophic collision between two larger bodies. But the Karin asteroid cluster discovered by Nesvorný and colleagues¹ is only a few million years old and may greatly improve our understanding of collisional dynamics and planet formation.

rely on much smaller laboratory experiments as well as numerical simulations to predict collision outcomes^{2,3}. But nature has provided a larger laboratory for us in the form of the ancient record of past collisions between asteroids. By studying this record, we can learn about the collision process.

One way to understand the physics of large-scale impacts among asteroids is to investigate asteroid families, groups of asteroids that share similar orbital elements and spectral properties⁴. Orbital elements describe the shape and orientation of the path that an object follows around the Sun, whereas spectral properties provide information about the composition of an object according to how it reflects sunlight at different wavelengths.

Asteroid families are formed during catastrophic collisions, as the fragments are launched away from the impact site at high speeds⁵. The nature of each break-up event and the velocity distribution of the ejecta depend critically on the incoming projectile's mass and velocity relative to the collision target. The ejection speed of the fragments acts as an impulse that gives each member a slightly different orbit from the parent body. In the absence of any

further disturbance, these orbits would remain unchanged.

In reality, however, these objects are subject to dynamical evolution through several mechanisms: subsequent collisions between family members and other nearby asteroids, after the family has formed; gravitational perturbations produced by planets and large asteroids; and non-gravitational forces such as the Yarkovsky effect that slowly cause the orbits of smaller, kilometre-sized asteroids to change over time⁶. (The Yarkovsky effect changes the spin and orbit of a body by the asymmetric re-radiation of thermal energy absorbed from the sun⁷.) Together, these effects slowly muddle the memory of the family-forming impact, making it difficult both to identify the family and to piece together exactly what happened. In fact, the degree of dynamical diffusion can be used to estimate how long ago the original impact occurred, although it is not a precise measure.

Until now, most known asteroid families were thought to be the by-products of huge disruption events among bodies hundreds of kilometres in diameter⁸. Most of these families are estimated to be hundreds of millions, even billions, of years old (Fig. 1), so it is problematic to use their properties to constrain asteroid evolution models. Using a new database of orbital elements, however, Nesvorný and colleagues¹ have identified a cluster of 39 small asteroids inside the Koronis asteroid family that were probably produced by the recent disruption of a 25-km asteroid.

The youth of this cluster is suggested by two lines of evidence. First, the orbital elements of the cluster members are remarkably similar. Second, the researchers were able to determine the precise age of the break-up event by numerically integrating the orbital elements of 13 cluster members back through time. They found that particular elements (corresponding to the orbital orientation of each body) converged to a single value 5.8 ± 0.2 million years ago. Because this technique does not work on asteroid clusters that have suffered significant dynamical evolution, this is the first time that an asteroid break-up has been accurately dated.

The Karin cluster — as Nesvorný *et al.* have named it, after its largest member — is a compelling target for a space mission. The cluster is young enough that many erosional and weathering processes thought to occur on asteroid surfaces⁹ may not have had time to

GALILEO/JPL/NASA

erase the tell-tale signatures of the break-up. So it might be possible to determine whether the family members are intact fragments or gravitational re-accumulations of smaller pieces. This new cluster will no doubt be the focus of attention for the asteroid community for some time. Meanwhile, the search for ever younger families will continue, in the hope of taking us closer to understanding the origins of our Solar System. ■

Derek C. Richardson is in the Department of Astronomy, University of Maryland at College Park, College Park, Maryland 20742, USA.
e-mail: dcr@astro.umd.edu

1. Nesvorný, D., Bottke, W. F. Jr, Dones, L. & Levison, H. F. *Nature* **417**, 720–722 (2002).
2. Housen, K. R., Holsapple, K. A. & Voss, M. E. *Nature* **402**, 155–157 (1999).
3. Asphaug, E., Ostro, S. J., Hudson, R. S., Scheeres, D. J. & Benz, W. *Nature* **393**, 437–440 (1998).
4. Zappalà, V., Bendjoya, Ph., Cellino, A., Farinella, P. & Froeschle, C. *Icarus* **116**, 291–314 (1995).
5. Michel, P., Benz, W., Tanga, P. & Richardson, D. C. *Science* **294**, 1696–1700 (2001).
6. Bottke, W. F. Jr, Vokrouhlický, D., Brož, M., Nesvorný, D. & Morbidelli, A. *Science* **294**, 1693–1696 (2001).
7. Farinella, P., Vokrouhlický, D. & Hartmann, W. K. *Icarus* **132**, 378–387 (1998).
8. Durda, D. D., Greenberg, R. & Jedicke, R. *Icarus* **135**, 431–440 (1998).
9. Clark, B. E. *et al.* *Meteor. Planet. Sci.* **36**, 1617–1637 (2001).

Ecology

Density and diversity

Hans ter Steege and Roderick Zagat

One explanation for the especially rich diversity of trees in the tropics is that a process called ‘density-dependent mortality’ operates there. It turns out, however, that this process occurs in temperate forests too.

The latitudinal gradient in species diversity is arguably the most universal pattern in global biodiversity: the lower the latitude, the higher the number of species in a given area. This pattern, with biodiversity peaking in the tropics (Fig. 1), is found in most taxonomic groups and may be as old as life itself^{1–3}. For plants, one explanation centres on a phenomenon called ‘density-dependent mortality’, in which the survival rates of species decrease as they become more common, leaving space for rarer species. It has been suggested^{4,5} that density-dependent mortality is more intense at lower latitudes, so, at least in part, accounting for the gradient in diversity. As they describe on page 732 of this issue, however, Hille Ris Lambers and colleagues⁶ have tested this hypothesis and found it wanting.

It is not surprising that many researchers

have sought to find an explanation for the latitudinal pattern in species diversity. Biologists study diversity at different scales and it is becoming clear that scale is an important bridging element between the various theories that have been developed for regional and local levels. As far as trees are concerned, local processes such as the rate and extent at which gaps for seedling colonization occur, as well as density-dependent mortality, may limit the extent to which one species excludes another, and thus promote local diversity. Density-dependent mortality, however, will maintain high diversity only if it is species specific — that is, if it decreases the density of a species as a function of the density of that species alone, rather than of the density of all species.

If species-specific, density-dependent mortality contributes to the latitudinal gradient in species diversity, the effect should be

stronger in the tropics than in temperate areas. The rationale for this is that much seed and seedling destruction stems from attack by insects and fungi, which reach higher densities in tropical regions because the tropics do not experience seasonal variations and tend to remain hot and humid^{4,5}.

Hille Ris Lambers *et al.*⁶, however, show that density-dependent mortality acts in temperate as well as in tropical forests. Their results come from a forest in North Carolina, where they found that six out of seven tree species experience density-dependent mortality at one or more transitions in their early stages: from seed to seedbank; from seed or seedbank to seedling; and in seedling survival (the seedbank phase is a latent period in which seeds lie dormant in the soil before germination). The results clearly show not just the effect of general density-dependent mortality, but also species-specific, density-dependent mortality. This suggests that rather than the driving factors being resource competition for light or nutrients alone, predators or pathogens are responsible. Hille Ris Lambers *et al.* also compared previous studies carried out in temperate and tropical areas, in which several species were tested for density-dependent mortality, and conclude that the proportion of species affected is not greater in the tropics.

All in all, this study⁶ adds support to those who claim that general ecological mechanisms, such as the creation of gaps in vegetation that allow colonization by seedlings, and density-dependent regulation, operate similarly in the tropics and temperate zones. So theories invoking these processes fail to explain the higher diversity in the tropics compared with temperate zones.

Still, further investigations are required. As Hille Ris Lambers *et al.* point out, many different protocols have been used in this kind of research, producing varying and sometimes confusing results. By re-analysing their own data with the approaches used in other studies, the authors show that these approaches have invariably underestimated the effects of density dependence. They propose a framework for simultaneously assessing the impact of seedling and adult density on seed and seedling fate. This framework should now be applied in future work at various latitudes.

The authors’ main message, then, is that the proportion of species subject to density-dependent regulation is not lower in temperate areas than in the tropics. But the ecologically more significant question may be to what extent species are actually regulated by the process. Are the effects strong or weak? And is there a latitudinal component? Addressing these questions will require identifying and quantifying latitudinal trends in the relevant factors. Earlier work demonstrated that density-dependent patterns of seed loss tend to be associated with predation by



Figure 1 Trees in the tropics — the height of diversity.

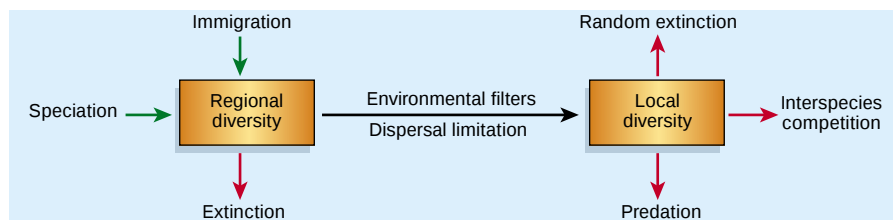


Figure 2 Determinants of biodiversity. Impacts on diversity occur at both regional and local scales^{8,9}; here, green arrows indicate processes that increase diversity (species are added) and red arrows those that decrease it (species are lost). Regionally, diversity is mainly influenced by speciation and extinction and, to a certain extent, by immigration. These processes determine the number of regional species from which communities can 'draw' particular species. Of course not all species can occur everywhere, either because the habitat is unsuitable (environmental filters) or because of constraints on their movement (dispersal limitation). Locally, processes such as interspecies competition, random extinction and predation affect diversity. Predation can have a double effect. It can remove species. But in the case of density-dependent mortality of seeds and seedlings discussed by Hille Ris Lambers *et al.*⁶, it can increase diversity if it favours rarer species by selectively reducing the population density of the most common species.

insects rather than by vertebrates, and that if spatial and temporal availability of seeds is predictable it can have an effect on seed loss⁷. So which predators determined the outcomes of the studies by Hille Ris Lambers *et al.*? And is there a latitudinal difference in the relative prevalence of insect and vertebrate predation, and of fungal attack? There is likely to be a difference in the fruiting time of particular plant communities, and in the occurrence of irregular heavy fruiting that swamps consumers and means a higher proportion of seeds escapes predation.

If indeed the small-scale processes observed in temperate and tropical areas are much the same, how can the latitudinal gradient in biodiversity be explained? Diversity is influenced by regional as well as local processes^{2,3,8}, all operating at specific temporal scales (Fig. 2). At the large spatial scale at which the gradient is evident, the answer must, at least partly, lie in the balance between speciation and extinction. A further question then centres on the relative importance of local processes in this balance. It is obvious that there are several mechanisms that can contribute to the latitudinal

gradient^{1-4,8}. There is unlikely to be a simple answer, and any explanation probably consists of a mix of local, regional, global and historical factors.

Hans ter Steege is at the International Institute for Geoinformation Science and Earth Observation (ITC), PO Box 6, NL-7500 AA Enschede, The Netherlands.

e-mail: tersteeg@itc.nl

Roderick Zagt is at the National Herbarium of the Netherlands, Utrecht University Branch, W. C. van Unnik Building, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands.

e-mail: lipaugus@luna.nl

1. Gaston, K. J. *Nature* **504**, 220–227 (2000).
2. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
3. Huston, M. *Biological Diversity: The Coexistence of Species on Changing Landscapes* (Cambridge Univ. Press, 1994).
4. Givnish, T. J. *J. Ecol.* **87**, 193–210 (1999).
5. Harms, K. E. *et al.* *Nature* **404**, 493–495 (2000).
6. Hille Ris Lambers, J., Clark, J. S. & Beckage, B. *Nature* **417**, 732–735 (2002).
7. Hammond, D. S. & Brown, V. K. in *Dynamics of Tropical Communities* (eds Newbery, D. N., Prins, H. H. T. & Brown, V. K.) 51–78 (Blackwell Science, London, 1998).
8. Ricklefs, R. E. & Schluter, D. *Species Diversity in Ecological Communities* (Univ. Chicago Press, 1993).
9. Hubbell, S. P. *The Unified Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).

Cardiovascular biology

A cholesterol tether

Bart Staels

The build-up of cholesterol in the walls of arteries is a hallmark of atherosclerosis. Work with transgenic mice has revealed a specific interaction through which cholesterol deposition is initiated.

Atherosclerosis is a disease of the arterial wall that is characterized by cholesterol accumulation and culminates in potentially life-threatening conditions such as heart attack, stroke and angina. How the process starts is poorly understood. In their paper on page 750 of this issue, however, Skålén and colleagues¹ show that at least

part of the explanation lies in a highly specific interaction between proteoglycans, which are components of the artery wall, and a constituent of lipoproteins called apoB.

The atherogenic process starts at an early age with the deposition in blood-vessel walls of lipids such as cholesterol, derived from lipoproteins circulating in the bloodstream,

which leads to the formation of the characteristic 'fatty streaks'. Inflammatory white blood cells congregate at these damaged areas through their interaction with adhesion molecules expressed by cells in the endothelial layer, which lines the inside of blood vessels. This can then set off a cascade of inflammatory processes and further lipid deposition, leading eventually to full-blown atherosclerosis with plaque formation in the artery wall. Atherosclerosis is thus viewed as a chronic inflammatory disease of the blood-vessel wall^{2,3}. An early event — the development of arterial damage and fatty deposits — is the capture, by cells called macrophages, of lipoproteins retained in the arterial wall, resulting in the formation of lipid-laden 'foam cells'. Although it has not been shown experimentally, the 'response-to-retention' hypothesis proposes that the retention and modification of these lipoproteins in the arterial lining is one of the initiating events that triggers the inflammatory response⁴.

Among the many risk factors for cardiovascular disease are raised concentrations of blood cholesterol, especially cholesterol transported by low-density lipoproteins (LDLs) containing the protein apoB100 (an apolipoprotein). Results from clinical trials in which concentrations of cholesterol in the blood stream were pharmacologically lowered established LDL cholesterol unequivocally as a culprit in atherosclerosis. But what triggers the accumulation of cholesterol-containing lipoproteins in the sub-endothelial space of the arterial wall? As long ago as 1949, Mogen Faber⁵ suggested that proteoglycans are involved — these are components of the extracellular matrix that are produced, for instance, by smooth-muscle cells that have migrated into the plaque. But because of a lack of suitable animal models, the effects of an interaction between proteoglycans and lipoproteins on atherogenesis have not been amenable to experimental testing *in vivo* — until now.

Skålén *et al.*¹ have analysed the susceptibility to atherosclerosis of transgenic mice carrying human apoB100 genes with a range of mutations that yield proteins with reduced affinity for proteoglycans⁶. They show that, when fed a diet that promotes atherogenesis, these mice develop plaques more slowly than control mice expressing the normal human apoB100 protein. The LDL particles from transgenic mice carrying normal and mutated human apoB100 were transported similarly across the endothelial lining of the wall and showed the same susceptibility to oxidation, inflammatory properties and capture by macrophages. So Skålén *et al.* conclude that the retention of apoB100 lipoproteins by proteoglycans immediately beneath the endothelial layer is a central event in early atherogenesis.

The strengths, as well as the limitations, of this study lie in the fact that Skålén and

colleagues tackled only one particular facet of atherosclerosis. In reality, many factors — both genetic and environmental — are likely to contribute to the disease. Not least, it can occur in the absence of raised levels of LDL cholesterol. A crucial question, then, is whether the interaction between apoB100 and proteoglycan is a general initiating mechanism for atherosclerosis, and it would be instructive to test its contribution to atherosclerosis induced by alterations in immune and inflammatory status. One approach would be to compare the susceptibility to atherosclerosis of normal and mutant human apoB100 transgenic mice cross-bred with mouse models of atherosclerosis caused by inflammation⁷.

Skålén *et al.*¹ also show that other apolipoproteins, such as apoE, which is found in triglyceride-rich and remnant lipoproteins, can substitute for apoB100 in mediating the retention of lipoproteins in the arterial wall. Increased levels of triglyceride-rich remnant lipoproteins are often found in the serum of patients suffering from insulin resistance and type II diabetes, as well as in people carrying particular variants of apoE. In such individuals, it could be that the interaction between apoE and proteoglycans is the greater contributor to the initiation of atherosclerosis.

The findings of Skålén *et al.* illuminate one molecular basis for lipoprotein retention in the arterial wall. But other mechanisms of atherogenesis — which might or might not depend on lipoprotein retention — could also be operating at the same time. Even given similar levels of LDL cholesterol, not every person who develops atherosclerosis does so to the same extent. Moreover, in a given individual, atherosclerotic plaques do not develop at an equal rate and are not evenly distributed throughout the vasculature. Contributing factors might include regional disturbances of blood flow or alterations in proteoglycan structure. The raised LDL concentrations due to overexpression of apoB100 in Skålén and colleagues' transgenic mice might induce an inflammatory response, resulting in increased proteoglycan expression and congregation of white blood cells at susceptible locations. Local inflammation at such sites would intensify the modification of LDL by activated cells of the arterial wall.

Whether lipoprotein retention and modification are involved in the ensuing course of atherosclerosis also remains an open question. Atherosclerotic plaques did form in mice with proteoglycan-binding-deficient human apoB100, albeit later than in mice expressing the normal form, which would seem to show that factors other than direct LDL retention by proteoglycans dominate after atherogenesis has been initiated. A major contributory factor may be lipoprotein lipase, an enzyme that is secreted by

macrophages in the artery wall and provides a high-affinity bridge between lipoproteins and proteoglycans⁸. It will be interesting to compare plaque development in normal and mutant apoB100 transgenic mice over a longer period. And are human subjects with mutations in the proteoglycan-interacting domain of apoB protected from atherosclerosis? It could be productive to search for these mutations in healthy octogenarians.

Even though it is clear that decreasing the concentration of cholesterol reduces the incidence and degree of atherosclerosis, most patients suffering from the disease still die from cardiovascular complications. Therapies that act directly on the arterial wall are needed, and Skålén and colleagues' results point to potential targets. One might consider, for example, using proteoglycan mimics that interfere with apoB100 retention in the wall. But great care and precision would be needed. Proteoglycans have other physiological functions that might be compromised by such an approach, and different proteoglycans might have different effects on

lipoprotein retention and thus in atherogenesis⁹. A complementary strategy would involve modulating the ability of LDLs to interact with proteoglycans by altering their composition and particle size. Given these new findings, measurements of the ability of LDL to bind to proteoglycans might produce a helpful diagnostic test, and decreasing the strength of the interaction would be a reasonable therapeutic approach. ■

Bart Staels is in the Faculté de Pharmacie, Université de Lille II, and UR545 INSERM, Département d'Athérosclérose, Institut Pasteur de Lille, 1 rue du Prof. Calmette, BP245, 59019 Lille, France.
e-mail: Bart.Staels@pasteur-lille.fr

1. Skålén, K. *et al.* *Nature* **417**, 750–754 (2002).
2. Ross, R. *N. Engl. J. Med.* **340**, 115–126 (1999).
3. Lusis, A. J. *Nature* **407**, 233–241 (2000).
4. Williams, K. J. & Tabas, I. *Arterioscler. Thromb. Vasc. Biol.* **15**, 551–561 (1995).
5. Faber, M. *Arch. Pathol. Lab. Med.* **48**, 342–350 (1949).
6. Boren, X. *et al.* *J. Clin. Invest.* **101**, 2658–2664 (1998).
7. Mallat, Z. *et al.* *Circ. Res.* **85**, 17e–24e (1999).
8. Pentikainen, M. O. *et al.* *Arterioscler. Thromb. Vasc. Biol.* **22**, 211–217 (2002).
9. Williams, K. J. *Curr. Opin. Lipidol.* **12**, 477–487 (2001).

Nanotechnology

Electronics and the single atom

Silvano De Franceschi and Leo Kouwenhoven

The invention of semiconductor transistors in the 1940s revolutionized electronic circuitry. In the new world of 'nanoelectronics', a transistor whose active component is a single atom has now been demonstrated.

Nanotechnologists are seeking to build nanometre-scale electronic devices in which the functional unit is a single molecule or atom¹. Carbon nanotubes have proved particularly useful as molecular 'wires'² in this quest, thanks to their long lengths (by nanotechnology standards at least) of several micrometres. So could it be possible to wire up a short molecule, or even a single atom, to create a nanoscale transistor? The experiments reported by Park *et al.*³ and Liang *et al.*⁴ (on page 722 and page 725 of this issue) give a positive answer.

Trapping a molecule between two metal electrodes to make such a transistor is a tough technological challenge^{1,3–7}. First, the molecule needs suitable terminations that reliably bind it chemically to the two electrodes, bridging the gap between them. Conventional lithographic techniques by which the electrode structure might be assembled have a resolution, at best, of around 10 nm — yet the electrodes would need to be just 1 nm apart. The two groups^{3,4} have used the unconventional combination of electron-beam lithography and electromigration to reach this 1-nm scale. Still, the entrapment of a single molecule in the electrode gap is an occasional, lucky event. Moreover, at present there is no viable imaging technique for

directly confirming successful trapping. Yet the presence of one, and only one, molecule can be indirectly established from its conduction properties.

The molecules used by Park *et al.* and Liang *et al.* in their ultra-small transistors are organic complexes that contain one³ or two⁴ atoms of a transition metal. The metal atoms, cobalt³ and di-vanadium⁴, form the active region of the device, whereas the organic molecule serves as mechanical support and provides the connection to the metal electrodes. Both experiments demonstrate transistor operation based on a tunable flow of electrons through the metal atom. (Although this simple picture is largely correct, the properties of the metal atom are strongly affected by the presence of the organic molecule.)

The current through an electronic transistor can be turned on and off by changing the voltage on a gate electrode. In the 'on' state, current is carried by a large number of electrons, with typically a billion passing per second — the large number is necessary to obtain a measurable current. In a commercial silicon transistor, electrons move independently of each other by diffusive motion from the 'source' to the 'drain' terminal. Although the motion of an individual electron cannot

be predicted, the average motion of a large number of electrons can, resulting in well-defined transistor operation.

How do electrons flow through a single-molecule transistor? The answer is by a simple fundamental process. The repulsive Coulomb interaction between electrons means that there is an energy cost in adding an extra electron to any object of small dimensions (this same energy cost is in part responsible for the ionization and affinity energies in an atom). In a molecular transistor, this energy cost can be tuned to zero by applying a voltage to the gate electrode. At a particular gate voltage, the electrostatic potential is such that an extra electron can hop from the source onto the molecule. However, Coulomb repulsion forbids a second, extra, electron hopping on at the same time — the first electron must leave the molecule, moving into the drain, to make way for the next electron. This one-by-one electron motion, governed by the quantum of electron charge, is known as single-electron tunnelling.

But there is another quantum property important for electron transport through small objects — the electron spin. Electrons each carry one half-unit of spin in one of two configurations, defined as 'up' or 'down'. As

there is no preference for the direction of the spin, the lowest-energy (ground) state of the system is a combination of up and down spins. To reach this ground state, an electron spin must flip between up and down, and this can be arranged by replacing an up-spin electron on the molecule by a down-spin electron from one of the electrodes. This spin-flipping, driven by the desire of the system to be in its ground state, enforces an exchange of electrons between molecule and electrodes through a process called the Kondo effect⁸.

In the experiments by Park *et al.*³ and Liang *et al.*⁴, the molecule is tuned between two states that differ by one charge and one spin unit. In one state, the conduction mechanism is single-electron tunnelling. In the next state, with an odd number of electrons on the molecule, electron transport is mediated by the Kondo effect. Remarkably, when the Kondo effect dominates, the conductance reaches values close to the quantum limit (which is $2e^2/h$, where e is the electron charge and h is Planck's constant) for a perfectly conducting one-dimensional wire. This result is quite remarkable, considering the intrinsic difficulty of establishing good electrical connection between a molecule and a metal electrode.

Do these realizations of a single-atom transistor mean that molecular electronics is just around the corner? That goal may be a little closer, but there is still a long road ahead before atomic or molecular transistors can be assembled into viable, dense, fast logic-circuits. Right now, these single-molecule or single-atom transistors are no competition for silicon transistors. But they will serve both scientifically, for studying electron motion through nanoscale objects, and technologically, for developing chemical techniques with which to fabricate electronic devices on single molecules.

Silvano De Franceschi and Leo Kouwenhoven are in the ERATO Mesoscopic Correlations Project, Department of NanoScience, Delft University of Technology, PO Box 5046, 2600 GA Delft, The Netherlands.

e-mail: leo@qt.tn.tudelft.nl

1. Joachim, C., Gimzewski, J. K. & Aviram, A. *Nature* **408**, 541–548 (2000).
2. Dekker, C. *Physics Today* **52**, 22–28 (1999).
3. Park, J. *et al.* *Nature* **417**, 722–725 (2002).
4. Liang, W., Shores, M. P., Bockrath, M., Long, J. R. & Park, H. *Nature* **417**, 725–729 (2002).
5. Park, H. *et al.* *Nature* **407**, 57–60 (2000).
6. Reichert, J. *et al.* *Phys. Rev. Lett.* **88**, 176804 (2002).
7. Zhitenev, N. B., Meng, H. & Bao, Z. *Phys. Rev. Lett.* **88**, 226801 (2002).
8. Kouwenhoven, L. & Glazman, L. *Physics World* **14**, 33–37 (2001).

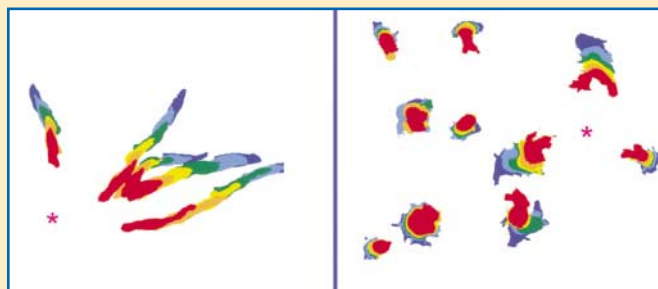
Cell motility

The attraction of lipids

The pictures reproduced here aren't just pretty — they reveal something fundamental about how cells move towards the source of an attractant chemical. The left image shows the behaviour of normal cells of *Dictyostelium discoideum*, a type of slime mould. The cells on the right lack an enzyme called PTEN. The colour scale, from blue to red, indicates how each cell moves over time. The normal cells are clearly moving towards a chemical source (asterisk), but the mutant cells seem less sure where they're going.

The images are taken from a paper by Miho Iijima and Peter Devreotes (*Cell* **109**, 599–610; 2002). Together with work by Satoru Funamoto *et al.* (*Cell* **109**, 611–623; 2002), this study reveals the importance to motility of enzymes — including PTEN — that alter the lipid content of cell membranes.

Many cells need to move up a concentration gradient of a given chemical ('chemoattractant'); in humans, for instance, immune cells migrate towards signals emitted by invading microorganisms. Cells



must sense the gradient and then extend protrusions ('pseudopodia') in the direction of movement. Filaments of the protein actin form the skeleton of pseudopodia.

Studies with lipid-binding proteins suggested that the lipid content of cell membranes forms a gradient during chemoattraction, with 3-phosphoinositides being concentrated at the front edge. These are thought to bind signalling proteins, leading to the changes needed for movement. But how does the lipid imbalance come about? This question is tackled in the new papers.

Funamoto *et al.* started by studying phosphatidylinositol-3-OH kinases (PI(3)Ks), enzymes that

produce 3-phosphoinositides, in *D. discoideum*. When cells were placed in a chemoattractant gradient, two different PI(3)Ks moved to the membrane at the leading edge, with similar kinetics to 3-phosphoinositide-binding proteins. Membrane-bound PI(3)Ks seemed to trigger the formation of pseudopodia.

So the localization of PI(3)Ks could explain the generation of 3-phosphoinositides at the leading edge. But the concentration of these lipids tails off sharply towards the rear of a cell, hinting that an enzyme that degrades them might be at work there. Both groups investigated a possible candidate —

PTEN. When Funamoto *et al.* overexpressed this enzyme in *D. discoideum*, the cells moved more slowly and became less polarized in response to a chemoattractant. Iijima and Devreotes engineered *D. discoideum* lacking PTEN, and again the cells moved abnormally, taking a circuitous route to the chemoattractant. The cells also produced pseudopodia more erratically, sometimes at the back, and had more actin filaments than normal. Finally, both groups found that PTEN is usually located at the rear of moving cells.

This suggests that PTEN and PI(3)Ks help to produce a gradient of 3-phosphoinositides, which is much steeper than the chemoattractant gradient. The lipid gradient translates into the correct localization of key signalling proteins and actin filaments, and so into directed movement. But how these enzymes are moved about the cell, and how they interact with other chemoattractant-sensing pathways, remains to be seen. **Amanda Tromans**

Physiology

Haemoglobin's chaperone

Lucio Luzzatto and Rosario Notaro

Molecular chaperones come in different forms, but all have a similar task: to keep other proteins in shape. A newly identified chaperone seems to be specific to haemoglobin, preventing precipitation.

The rate of production of consumer goods is controlled at the assembly line, but it must be also adjusted according to demand. In biological terms, the synthesis of haemoglobin is an extreme example of what happens when market demands are high. Haemoglobin is the protein that makes red blood cells red. It is responsible for delivering oxygen to all tissues and organs in the body, and to do so optimally it accumulates in red blood cells to an incredible 340 grams per litre¹. In adults, functional adult haemoglobin (HbA) consists of two α - and two β -globin chains, each containing haem — a non-protein group that carries oxygen². The two chains cooperate to ensure that HbA can bind and release oxygen in an efficient and well-controlled way.

The control of globin gene expression is complex^{3,4}, and the need for two different globin chains makes the assembly line even more complicated, especially because the two genes are on different human chromosomes^{5,6}. The end result is a high rate of haemoglobin production, with absolute tissue specificity. A long-standing question⁷ is how red blood cells ensure that both globin genes are top performers and yet provide

their protein products in perfect stoichiometry. On page 758 of this issue, Kihm and colleagues⁸ provide at least part of the answer.

Among several differences between the α - and β -globin chains, two are especially important. At the genomic level, there are two α -globin genes to every β -globin gene. At the protein level, the β -chain can associate on its own with haem and form a tetramer, called HbH. This tetramer is functionally useless because, although it does bind oxygen, it cannot easily release it. By contrast, haem-bound α -chains on their own tend to form precipitates, called α -inclusion bodies, that damage red blood cells⁹.

We can now perceive how stoichiometry might work. If red blood cells produced β -chains in excess, there would be wasteful synthesis of HbH. So one can imagine that it would be advantageous for α -chains to be produced in slight excess, and there is experimental evidence that this does indeed happen. Thus, α -chains will combine with every available β -chain to form functional HbA. But this arrangement will work only if the intrinsic instability of the excess α -chains does not cause harm to the cell or cell membranes. Kihm *et al.*⁸ now show that a

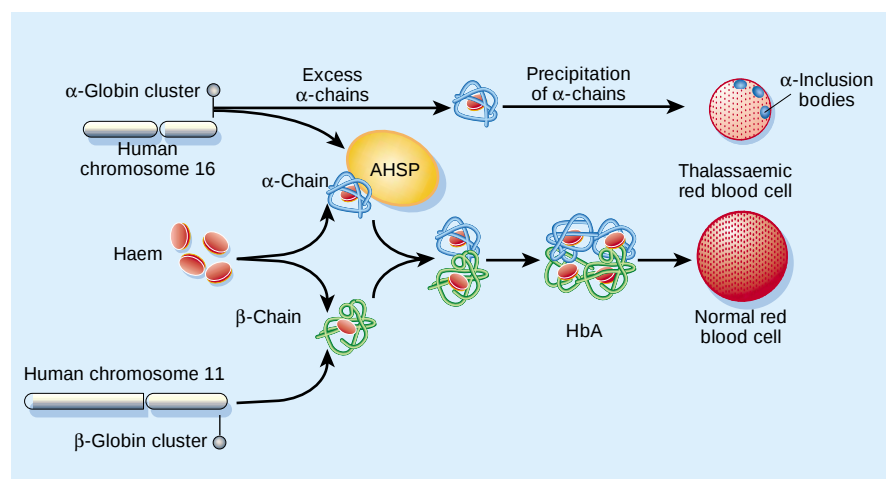


Figure 1 Balancing the components of haemoglobin. Left, the α -globin and β -globin chains are encoded by genes on different chromosomes and so their expression (which is specific to red blood cells) is controlled independently. Slightly more α -chains than β -chains are produced (not shown). Having incorporated the oxygen-carrying haem group, the α - and β -chains cooperate to form first dimers and then tetramers (lower pathway). Kihm *et al.*⁸ find that AHSP binds specifically to α -haemoglobin, and suggest that it might thereby serve a dual purpose — to stabilize these chains and to help in delivering newly formed α -chains to β -chains. Top pathway, in β -thalassaemia fewer β -chains are produced and the excess in α -chains is too great for AHSP to handle; some α -chains precipitate, damaging the cell. The involvement of AHSP in the production of haemoglobin is corroborated by the finding⁸ that red blood cells from AHSP-deficient mice have features reminiscent of thalassaemia.



100 YEARS AGO

We should like to draw your attention to the following spectacle which some of us witnessed on the sea-shore at Blundellsands on Thursday evening, June 5, at about eight o'clock. The evening was dull and grey, a strong north-westerly wind was blowing in from the sea and the tide was flowing in. In the distance we first saw smoke with frequent jets of fire bursting forth from the mud of a shallow channel. Drawing near, we perceived a strong sulphurous odour, and saw little flames of fire and heard a hissing sound as though a large quantity of phosphorus was being ignited. It was impossible to detect anything which caused the fire... The area over which the tiny flames kept bursting forth was about 40 yards. A gentleman present stirred up the mud with his walking-stick, and immediately large yellow flames nearly 2 feet in length and breadth burst forth. The phenomenon lasted some time, until the tide covered the part and quenched the fire. From *Nature* 12 June 1902.

50 YEARS AGO

In Sweden, as in Britain, observations have been made of the opening of milk bottles by blue tits and great tits... Although no experimental analysis of the behaviour involved in the opening of milk bottles has yet been made, further observations in the field enable the discussion to be carried further. Previously the processes were considered in two parts — the 'recognition' of the milk bottle as a potential supply of food and the technique of opening the bottle. The actual opening of the bottle probably depends on innate motor patterns, and the study of tits in the field has confirmed this view... It has been suggested that the initial 'discovery' of the bottle as a source of food may be a logical consequence of the feeding habits of tits... In fact, it seems that, when tits are looking for food, objects with certain very general characteristics may be sufficient to release a more limited type of searching behaviour (such as flying to the food). By a succession of similar steps (hammering, etc.), evoked by successively more specific stimuli, a reward may eventually be obtained. It so happens that milk bottles have sufficient characteristics in common with the natural foods for tits occasionally to 'discover' them in the course of their normal searching. From *Nature* 14 June 1952.

Daedalus

Collimated gas

A hot gas, says Daedalus, is isotropic. You get the same temperature no matter how you hold the thermometer. If you want the gas to give some anisotropic push, you need a mechanical element — the rocket nozzle and the turbine blade are probably the best known. DREADCO chemists now want to generate anisotropic heat directly. Any fluid becomes isotropic with time, so fast reactions, explosions or rocket-reactions, seem most hopeful. Furthermore, all the energy is stored in the initial molecules. No entering oxygen complicates their anisotropic decomposition.

In this connection Daedalus recalls gunpowder, one of the cleverest of chemical inventions. At present it is made by simple mixing. The DREADCO team are devising a 'linear gunpowder' in which carbon fibres are laid alongside or interwoven with sulphur fibres (made by pouring molten sulphur into water), and these are interwoven with potassium nitrate whisker crystals. The resulting linear product, carefully aligned, should burn in the fibre-direction much faster than in the perpendicular one. Its gas will be hotter and moving faster in that direction, and colder at right-angles. The average temperature will, however, be as for normal gunpowder.

Even a small effect will be well worth having. In burning, linear gunpowder will eject an already-oriented gas. In a rocket, it will give added thrust. The rocket nozzle will have less to do, and will be burnt away more slowly by the cool sideways gas it feels. The DREADCO team also recalls how cordite is extruded as a gel in acetone — perhaps this too could be molecularly aligned anisotropically, thus gaining efficiency. The military would certainly value even a slight increase in the thrust of its rockets.

Of course, the biggest solid-fuel rockets are those of cold-war weapons and the Space Shuttle. Aligned and anisotropic fuels for these monsters could make them even more effective. On a smaller scale, the brass cartridge-case is now filled in bulk with isotropic propellant. Any ordinary gun would benefit from an oriented cartridge whose anisotropic propellant delivered its high-velocity molecules preferentially to the bullet and the breech, with less wasteful heating directed sideways at the barrel. And, of course, the delicate art of making fast- and slow-burning fuses for mining and quarrying would be greatly expanded.

David Jones

previously discovered protein, which they call α -haemoglobin-stabilizing protein (AHSP), seems to prevent just this problem (Fig. 1).

Kihm *et al.* started by screening for genes that are turned on by GATA-1, a gene-transcription factor that regulates the production of globins and of the enzymes required to synthesize haem¹⁰. The screen led to the identification of one gene, expressed at relatively high levels in red blood cells, whose encoded protein — AHSP — binds specifically to α -haemoglobin *in vitro*; it does not bind to β -haemoglobin or to HbA. Kihm *et al.* also show that when α -haemoglobin is artificially overexpressed in suitable test cells, it forms large precipitates. However, when α -haemoglobin and AHSP are expressed in the same cells, both proteins remain distributed homogeneously throughout the cytoplasm. Similarly, AHSP prevents free α -haemoglobin from precipitating *in vitro*, whether spontaneously or after the oxidation of haem. Finally, like a truly enlightened chaperone, AHSP keeps a bound α -haemoglobin under control only until the latter encounters the desired partner, a β -haemoglobin.

These results could have medical implications. It was realized long ago that the balance of the α - and β -globin chains is important not just to molecular physiology but also in blood disease. Indeed, in β -thalassaemia syndromes the main defect is by definition a substantial reduction in the rate of β -chain synthesis, leading to less HbA per red blood cell. When both of the two copies of the β -globin gene are defective (that is, in the homozygous state), patients suffer from severe anaemia, which they can survive only if they receive regular blood transfusions or a bone-marrow transplant from a suitable donor¹¹.

In these patients, because of the lack of β -chains there is a relative excess of α -chains¹², and this is a major determinant of the severity of disease¹³. For instance, people who have a β -thalassaemia-causing mutation in just one copy of the β -globin gene (they are heterozygotes) have essentially no symptoms. However, they will develop a relatively serious condition, thalassaemia intermedia, if they also have a triplicated α -globin gene¹⁴. Conversely, people who would normally develop severe β -thalassaemia (thalassaemia major) as a result of a homozygous β -globin mutation can have a milder condition, again thalassaemia intermedia, if they also have mutations in the α -globin gene that reduce the concentration of the α -chain¹⁵.

Could AHSP be relevant to these diseases? The last part of the paper by Kihm *et al.*⁸ suggests that the answer might be yes. The authors engineered mice that lacked functional AHSP, and found remarkable changes in the animals' blood. The mice did not have thalassaemia, because their globin genes were intact. But their red blood cells

showed abnormalities consistent with damage caused by unchaperoned α -chains. This provides *in vivo* support for the idea that the slight excess of α -chains in normal red blood cells is effectively neutralized by AHSP. By contrast, in homozygous β -thalassaemia patients, because there are no β -chains for the α -chains to pair up with, the α -chains will exceed the chaperone capacity of AHSP (the intracellular molar ratio of AHSP to α -chains is only about 1:50). The resulting damage would cause the death of maturing red blood cells.

Kihm *et al.* also speculate that naturally occurring AHSP mutations could modify the clinical picture of β -thalassaemia. There are at least two crucial tests of this idea. First, the offspring produced by mating AHSP-deficient mice with animals heterozygous for β -thalassaemia might not survive (much like mice with homozygous β -thalassaemia mutations¹⁶). Second, in humans, unexplained cases of thalassaemia intermedia in β -thalassaemia heterozygotes might result from mutations in AHSP that would cause the α -chain excess to be more detrimental.

The reverse situation is perhaps more difficult but is also more attractive to a haematologist. Could it be that there are mutations that result in the overexpression of AHSP, so converting thalassaemia major into thalassaemia intermedia? In every large thalassaemia centre there are rare patients, homozygous for a severe β -globin mutation, who do not depend on transfusions. Perhaps we should be studying the AHSP gene from these people first. If a mutated or overexpressed AHSP gene were so beneficial, one might be tempted to consider it for future gene therapy. In that way, a strict chaperone might become a loving nurse.

Lucio Luzzatto and Rosario Notaro are in the Laboratory of Human Molecular Genetics, IST, National Institute for Cancer Research, Largo Rosanna Benzi 10, 16132 Genova, Italy. e-mail: lucio.luzzatto@istge.it

1. Wintrobe, M. M. *J. Lab. Clin. Med.* **17**, 899 (1932).
2. Fermi, G. & Perutz, M. F. *Atlas of Molecular Structure in Biology: Haemoglobin and Myoglobin* (Clarendon, Oxford, 1981).
3. Cantor, A. B. & Orkin, S. H. *Curr. Opin. Genet. Dev.* **11**, 513–519 (2001).
4. Grosfeld, F. *et al. Phil. Trans. R. Soc. Lond. B* **339**, 183–191 (1993).
5. Van der Ploeg, L. H. *et al. Nature* **283**, 637–642 (1980).
6. Higgs, D. R., Sharpe, J. A. & Wood, W. G. *Semin. Hematol.* **35**, 93–104 (1998).
7. Baglioni, C. & Campana, T. *Eur. J. Biochem.* **2**, 480–492 (1967).
8. Kihm, A. J. *et al. Nature* **417**, 758–763 (2002).
9. Fessas, P. *Blood* **21**, 21 (1963).
10. Tsai, S. F. *et al. Nature* **339**, 446–451 (1989).
11. Weatherall, D. J. & Clegg, J. B. *Thalassaemia Syndromes* (Blackwell, Oxford, 2001).
12. Clegg, J. B., Naughton, M. A. & Weatherall, D. J. *J. Mol. Biol.* **91**, 91–108 (1966).
13. Nathan, D. G., Stosel, T. B., Gunn, R. B., Zarkowsky, H. S. & Laforet, M. T. *J. Clin. Invest.* **48**, 33–41 (1969).
14. Galanello, R. *et al. Blood* **62**, 1035–1040 (1983).
15. Kan, Y. W. & Nathan, D. G. *J. Clin. Invest.* **49**, 635–642 (1970).
16. Yang, B. *et al. Proc. Natl Acad. Sci. USA* **92**, 11608–11612 (1995).

Obituary

Stephen Jay Gould (1941–2002)

Stephen Jay Gould, the world's most renowned palaeontologist, died in New York on 20 May. His death robs the fields of palaeontology and evolution of one of their most provocative thinkers, and millions of readers of an entertaining and astonishingly productive commentator on biology.

Gould was born in New York City, and traced his lifelong interest in palaeontology to an encounter as a five-year-old with a dinosaur skeleton in the American Museum there. After obtaining a degree in geology from Antioch College in 1963, he became a graduate student at the museum in the joint programme with Columbia University. He acknowledged the importance of Norman Newell as his advisor and teacher ("the most noble word of all human speech"). In 1967 Gould was appointed assistant professor at Harvard University where he willingly took on the teaching of undergraduate non-majors. Even when he became a household name and was much in demand for public lectures, he continued to teach courses in geology, biology and the history of science at Harvard, where he spent his entire career.

Gould's style of science lecturing was unconventional, and he rejoiced in using illustrations and examples from history and art. He liked to read from historical sources, which he sometimes carried with him, and he could quote swathes of the *Origin of Species* from memory. He was an eloquent advocate for the importance and rigour of the historical sciences, and railed against the common perception of physics and chemistry as somehow superior in status. His digressions and allusions sometimes challenged the audience — bullet points were not Gould's style. But just as the journey was full of interesting detours, the destination was always worth reaching.

Gould's principal scientific contribution was to the theory of evolution, but he was no stranger to field data. His graduate research was on the evolution of the land snails of Bermuda, on which he continued to work throughout his career, notably in collaboration with geneticist David Woodruff. Gould first detected the phenomenon of morphological stasis punctuated by periods of rapid evolutionary change in the snail *Poecilozonites*. It was this pattern that he and Niles Eldredge, who identified the same pattern in the Devonian trilobite



Palaeontologist and public face of evolutionary biology

Phacops, named 'punctuated equilibrium' in 1972. They contrasted it with 'phyletic gradualism', the gradual transformation of one species into another.

Eldredge and Gould's model provoked a storm of debate, and, like much of Gould's later thinking, stimulated many other researchers, in this case to explore the fossil evidence for the nature of speciation. As Gould and Eldredge wrote in 1993, the "most important implications" of punctuated equilibrium "remain the recognition of stasis as a meaningful and predominant pattern within the history of species, and the recasting of macroevolution as the differential success of certain species (and their descendants) within clades". Thus Gould rejected the reductionist view, championed by Richard Dawkins, that natural selection operates exclusively on discrete genes, but also the idea that selection on individuals in populations is the sole force shaping large-scale evolutionary trends.

Gould's best-known book is probably *Wonderful Life* (1989), which recounts the story of the reinterpretation of the remarkable animals of the Burgess Shale of British Columbia by Harry Whittington and his students. These fossils are Cambrian in age (over 500 million years old), and Gould's thesis is simple. If the fish-like creature *Pikaia*, the earliest known chordate, had perished in some Cambrian extinction we, its descendants, would not exist. The evolutionary outcome is contingent on everything that came before. Replay the tape of life and today's biota looks completely different.

Contingency, Gould argued, is a major factor in macroevolution, not least because the response of organisms to catastrophes such as asteroid impacts cannot be predicted — the normal rules of natural selection do not apply. There is no inevitable evolutionary 'progress' to higher forms of life, culminating in man.

Two of Gould's earlier books are scholarly monographs: *Ontogeny and Phylogeny* (1977) considers the theory of recapitulation (the idea that the evolutionary history of an organism is reflected in the stages of its development) and the *Mismeasure of Man* (1981, 1996) treats measures of human intelligence and their past misuse in ranking racial groups. Between 1974 and 2001 Gould wrote 300 essays under the banner "This view of life" for the magazine *Natural History*. The tenth and final volume of these collected essays, entitled *I Have Landed*, was published earlier this year. He was a vigorous opponent of creationism, and particularly the movement to give equal time to creation teaching in schools. These articles are testimony to his commitment to making serious science accessible to a wider audience.

In 1982, Gould was diagnosed with mesothelioma, a rare cancer usually associated with exposure to asbestos. With characteristic intellectual curiosity, he consulted the medical literature and decided that he had a good chance of falling within the tail of the right-skewed distribution of survival times. During his treatment, he never failed to deliver his monthly columns for *Natural History* magazine. Sadly, 20 years later, he succumbed to a different cancer that had spread before it was detected.

Gould's magnum opus, *The Structure of Evolutionary Theory*, was published just a few months before he died. Over 20 years in preparation and more than 1,400 pages long, it traces the history of ideas on evolution. Gould considers the main challenges to Darwin: that evolution is hierarchical, operating not only on genes but on species; that natural selection is not the only engine of evolution; and that major perturbations — catastrophic events — influence the fate of groups. He then presents his own synthesis of evolutionary theory in a way that will provoke discussion and research for many years, and serve as a memorial to this remarkable man.

Derek E. G. Briggs

Derek E. G. Briggs, Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK, is currently on sabbatical in the Department of the Geophysical Sciences at the University of Chicago.
e-mail: D.E.G.Briggs@bristol.ac.uk

Solving the puzzle of mirror-image flowers

The genetically controlled orientation of floral sex organs encourages cross-pollination.

Enantiostyly is a plant sexual polymorphism in which female sex organs are deflected to the left or right — resulting in ‘mirror-image’ flowers — but, although it occurs in at least a dozen unrelated families of flowering plants¹ and has been known for over a century², its adaptive significance has been unclear. Here we show that a mendelian locus governs the inheritance of style orientation and that this curious form of sexual asymmetry functions to promote cross-pollination in bee-pollinated plants.

To investigate whether enantiostyly is genetically determined, we carried out controlled crosses on *Heteranthera multiflora*, a species with left- and right-styled plants. Progeny ratios in families of the first and second generations (F_1 and F_2) indicated that style deflection is governed by a mendelian locus at which the right-styled allele (R) is dominant to left (r).

All crosses between genotypically homozygous plants with opposite style orientations (8 families, $n=242$ progeny)

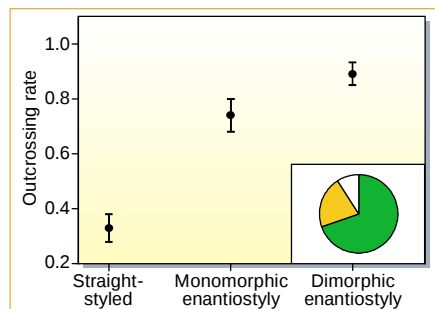


Figure 2 Enantiostyly influences mating patterns in experimental arrays of *Solanum rostratum*. Mean outcrossing rates were determined by using polymorphic allozyme loci⁶ for different style orientations in arrays of 16 plants (trimmed to 6 flowers per plant) in a square grid (4 replicates each). Straight-styled plants, the style was tied to the base of the pollinating anther with clear nylon thread; monomorphic enantiostyly, left- and right-styled flowers on the same plant; dimorphic enantiostyly, flowers were removed to create arrays that were entirely left- or right-styled. Outcrossing rates were estimated for all array types; the frequency of inter- and intramorph mating was determined for dimorphic arrays. Array type had no significant effect on the mean number of flowers visited per plant by bumble-bees (log-likelihood ratio, 4.11; $P=0.39$), which did not discriminate between left- and right-styled flowers (log-likelihood ratio, 2.46; $P=0.12$). Values are means $\pm$ s.e. based on 1,000 bootstrap estimates; ANOVA was used for outcrossing-rate differences, with the effect of each observation weighted by the inverse of its squared standard error to account for variation in the estimates⁶. Outcrossing rates differed significantly between the three treatments (ANOVA: $F=40.38$; d.f. = 2; $P<0.001$; *a priori* contrasts: straight-styled vs monomorphic enantiostylous, $F=21.43$; d.f. = 1; $P=0.001$; monomorphic vs dimorphic, $F=59.33$; d.f. = 1; $P<0.001$). Inset, frequency of intermorph (green), intramorph (orange) and selfed (white) mating events in dimorphic enantiostylous arrays.

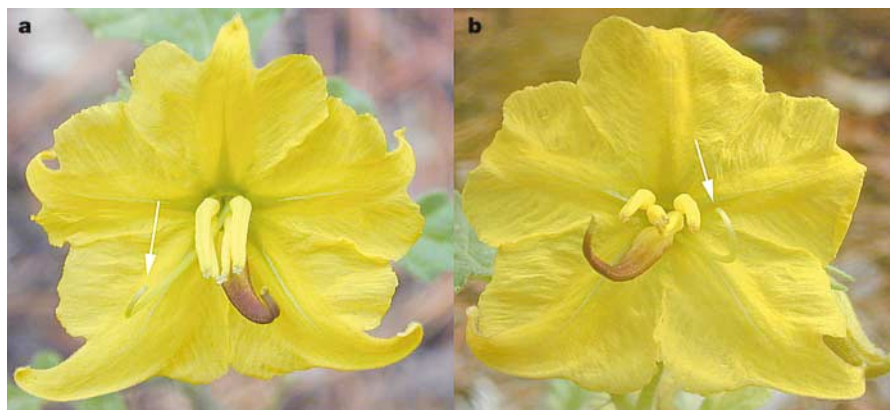


Figure 1 The ‘mirror-image’ flowers of *Solanum rostratum*. **a**, Left-styled flower; **b**, right-styled flower. Styles are indicated by arrows.

produced only right-styled plants. All F_2 segregations conformed to the expected 3:1 ratio of right- to left-styled plants (6 families, $n=242$ progeny; right, 0.76; left, 0.24; deviation from 3:1, $G_{\text{het}}=1.273$, $P=0.93$; $G_{\text{pooled}}=0.139$, $P=0.71$). The discovery of a mendelian locus that controls style orientation is, to our knowledge, the first demonstration of a gene for ‘handedness’ in plants.

With single-locus control of style orientation and equivalent levels of cross-pollination between left- and right-styled plants, a 1:1 ratio of the two style morphs should occur in natural populations of enantiostylous plants. This is the case in *Wachendorfia paniculata* populations³, suggesting that polymorphism in this species is also likely to be simply inherited.

Bees visiting enantiostylous flowers pick up pollen on the sides of their bodies, which favours transfer between flowers of opposite style deflection⁴ and suggests that enantiostyly promotes cross-pollination in a way that is functionally analogous to heterostylous sexual polymorphisms⁵. But although this makes sense for enantiostylous species in which plants have one or other stylar direction (dimorphic enantiostyly), species that produce both left- and right-styled flowers on an individual plant (monomorphic enantiostyly) are common and present a problem for the cross-pollination hypothesis¹. This is because the presence of left- and right-styled flowers on the same plant could promote interfloral self-pollination, resulting in inbreeding depression and loss of outcross mating opportunities through pollen discounting⁶.

However, monomorphic enantiostyly probably evolved from an ancestor with straight styles. In straight-styled plants, all flowers on the plant can potentially donate and receive pollen from one another, whereas between-flower transfer should be

relatively reduced in monomorphic enantiostyly (depending on the ratio of left- and right-styled flowers on a plant).

We investigated the role of enantiostyly in promoting cross-pollination in experimental arrays of *Solanum rostratum* (Fig. 1), a species with monomorphic enantiostyly that is visited primarily by bumblebees⁴. By using marker genes and manipulating the flowers and inflorescences, we compared mating patterns in arrays composed of plants with the same number of flowers, but with each array having one of three stylar arrangements: straight styles, monomorphic enantiostyly or dimorphic enantiostyly (Fig. 2).

We predicted that outcrossing rates should be highest for arrays with dimorphic enantiostyly, lower for those with monomorphic enantiostyly, and lowest in those composed of straight-styled plants. For dimorphic enantiostyly, most mating should also occur between plants with opposite style orientations. Our results confirmed these predictions (Fig. 2).

We have shown that enantiostyly can increase the precision of cross-pollen transfer. The multiple origin of mirror-image flowers among angiosperm families provides a striking example of a convergent floral strategy that promotes pollen dispersal in bee-pollinated plants.

Linley K. Jesson, Spencer C. H. Barrett
Department of Botany, University of Toronto,
25 Willcocks Street, Toronto,
Ontario M5S 3B2, Canada
e-mail: barrett@botany.utoronto.ca

- Barrett, S. C. H., Jesson, L. K. & Baker, A. M. *Ann. Bot.* **85** (suppl. A), 253–265 (2000).
- Todd, J. E. *Am. Nat.* **16**, 281–287 (1882).
- Jesson, L. K. & Barrett, S. C. H. *Am. J. Bot.* **89**, 253–262 (2002).
- Bowers, K. A. W. *Am. J. Bot.* **62**, 633–638 (1975).
- Barrett, S. C. H. *Nature Rev. Genet.* **3**, 274–284 (2002).
- Harder, L. D. & Barrett, S. C. H. *Nature* **373**, 512–515 (1995).

Competing financial interests: declared none.

Tropical agriculture

The value of bees to the coffee harvest

The self-pollinating African shrub *Coffea arabica*, a pillar of tropical agriculture, was considered to gain nothing from insect pollinators^{1,2}. But I show here that naturalized, non-native honeybees can augment pollination and boost crop yields by over 50%. These findings, together with world coffee-harvest statistics and results from field studies of organically shade-grown coffee, indicate that coffee plants would benefit from being grown in habitats that are suitable for sustaining valuable pollinators.

African honeybees colonized western Panama in 1985, where they naturalized. By 1997 they had become major pollinators of coffee growing near forests at 1,500 m above sea level². Yields of *C. arabica* may therefore be higher near forest, which provides a good pollinator habitat². In a study of 50 2-year-old plants in Panama in 2001, I observed that flowers were visited not only by native pollinators, but also by the naturalized honeybees.

Ripe berries resulting from open pollination of coffee flowers were heavier than those on control branches that had been bagged with fine-mesh material (from which pollinators were excluded), and were more abundant per flower (49% increase; $P < 0.01$, paired *t*-test). The open-pollinated fruit was, on average, 7% heavier, whereas a 25% increase in mass was recorded when African honeybees had exclusively dominated the flowers². This suggests that the contributions to final berry weight and total yield^{1,2} may differ for non-native honeybees and other, natural pollinators; however, bees consistently controlled over 36% of the total production.

Do bees control coffee harvests on a larger scale? Long-term data indicate that they do, although the results require detailed analysis. Almost 11 million hectares of coffee were harvested in 2001 (ref. 3). Cultivated areas of coffee in Ivory Coast, Ghana, Kenya, Cameroon and Indonesia have increased two- to fivefold in the past 41 years, although yields have decreased by 20–50%. El Salvador and Haiti, like other countries with intensive land usage and little natural habitat, show similar trends (Table 1)³. Pollinator loss is implicated in this decline, as sustained and aggressive cultivation may harm pollinators by removing their habitat.

A substantial increase in Latin American coffee yield partly coincided with the establishment of African honeybees in those countries^{2,4}, although there was no such change in the Old World, where honeybees originated (Table 1). This comparison under-

Table 1 Worldwide coffee yields and potential contribution by African honeybees

New World	1961–80	1981–2001	Old World	1961–80	1981–2001
Costa Rica	9,139; 0.21	14,620; 0.09	Vietnam*	4,251; 0.44	13,380; 0.37
Bolivia*	7,686; 0.09	8,767; 0.09	Papua NG*	10,522; 0.19	11,561; 0.21
El Salvador†	9,996; 0.11	8,729; 0.10	Thailand*	3,509; 0.83	9,652; 0.24
Guatemala	5,488; 0.16	8,231; 0.11	Philippines††	10,028; 0.20	9,593; 0.13
Colombia	5,920; 0.06	7,740; 0.16	Ethiopia	5,667; NA	8,797; NA
Honduras	4,096; 0.18	7,264; 0.14	Sri Lanka†	17,064; 0.24	7,869; 0.23
Nicaragua	4,566; 0.28	6,408; 0.23	India	5,904; 0.20	7,354; 0.21
Brazil	4,965; 0.25	6,283; 0.21	Malaysia	4,001; 0.19	7,288; 0.13
Peru	5,487; 0.09	5,740; 0.10	Kenya*	6,604; 0.15	6,055; 0.22
Mexico*	5,227; 0.12	5,116; 0.14	Tanzania	4,738; 0.11	5,432; 0.10
Haiti†	5,226; 0.04	5,024; 0.02	Indonesia†	5,716; 0.05	5,394; 0.08
Panama	2,347; 0.20	4,124; 0.14	Madagascar†	3,824; 0.08	3,576; 0.08
Dominican Republic	3,145; 0.12	3,949; 0.25	Cameroon††	3,525; 0.19	3,141; 0.21
Ecuador*	3,089; 0.14	3,240; 0.25	Ivory Coast*††	3,393; 0.25	2,391; 0.30
Venezuela	1,938; 0.15	2,789; 0.16	Ghana	3,213; 0.57	2,136; 0.53
Mean‡	5,380	6,850		6,130.6	6,907
Standard deviation	2,413	3,057		3,776	3,346
Paired <i>t</i> -test	0.004			0.232	

Mean average coffee yield² (in kg ha⁻¹) and coefficient of variation for different countries in the New and Old Worlds are shown for two periods; 1981–2001 coincides with the continent-wide presence of feral African honeybees in tropical America. Paired *t*-tests (one-tailed) evaluate changes in both New World and Old World yields. NA, not applicable.

*Aggressive high-density cultivation. †Negative exponential trend in yield³. ‡Caribbean islands excluded.

lines a possible cause-and-effect relationship between the presence of social bees and coffee yield.

Further comparison of yields from the Caribbean islands with those from Mexico and Central America suggests that social bee colonies, which exploit blooming coffee intensely, have two important effects. Such colonies, whether native or introduced, are virtually absent on Caribbean islands. On the mainland, African honeybees may replace native pollinators (primarily stingless, social bees) without affecting yields (paired *t*-test)^{2–4} but they reduce the variation in yield, as indicated by the coefficient of variance (c.v., ratio of the standard deviation to the mean).

The c.v. magnitude for the islands had been the lowest in the region by a factor of two and has been stable for 41 years ($P = 0.27$, paired *t*-test). But after African bees arrived in Central America and Mexico, it dropped for those areas ($P = 0.02$, paired *t*-test), eventually reaching a value that is 23% less than for Caribbean islands. Moreover, the coffee yield of the islands has remained only half that of Central America and Mexico³, indicating an absence of pollination and outcrossing benefits from bees². The low c.v. in yield in Haiti, for example (Table 1), may be derived from low variance in self-pollination and scarce pollinators.

Recent saturation of the neotropics with feral honeybees, which compete with other flower visitors⁴, has caused intensive exploitation of coffee and other flowering plants and has promoted pollination stability.

However, although the island of New Guinea has no honeybees, its yields remain high (Table 1), partly because its native solitary bees pollinate the obligately outcrossing coffee plant *Coffea canephora* there^{1,3}. *C. canephora* is grown in tropical lowlands and extensively in the Old World, but it is also wind-pollinated¹.

Declining yields can be offset by expanded cultivation or by increasing planting density, but such remedies are unstable (Table 1). Although shade conditions significantly improve the flavour of commercial coffee^{5,6}, coffee monocultures often lead to the removal of shade trees. The trend towards cultivating 'sun coffee' at high densities to boost yield^{5,6} will eliminate sites for bee nesting and mask the erosion of pollinator populations, which is shown here to affect yield by 36%. Optimization of coffee harvests and agricultural flexibility in tropical countries in the long term will depend on a consideration of pollinator sustainability and habitat.

David W. Roubik

Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002-0948, USA
e-mail: roubikd@tivoli.si.edu

- Free, J. B. *Insect Pollination of Crops* 2nd edn (Academic, London, 1993).
- Roubik, D. W. *Int. Workshop Conserv. Sust. Use Pollinators* (eds Kevan, P. G. & Imperatriz-Fonseca, V.) (Int. Bee Res. Assoc., Cardiff, in the press).
- FAOSTAT online; <http://apps.fao.org>
- Roubik, D. W. *Science* **201**, 1030–1032 (1978).
- Muschler, R. G. *Agroforest. Syst.* **51**, 131–139 (2001).
- Lyngbaek, A. E., Muschler, R. E. & Sinclair, F. L. *Agroforest. Syst.* **53**, 205–213 (2001).

Competing financial interests: declared none.

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com/nature) or on request from nature@nature.com

Architecture for a large-scale ion-trap quantum computer

D. Kielpinski^{*}, C. Monroe[†] & D. J. Wineland[‡]

^{*} Research Laboratory of Electronics and Center for Ultracold Atoms, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[†] FOCUS Center and Department of Physics, University of Michigan, Ann Arbor, Michigan 48109-1120, USA

[‡] Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado 80305, USA

Among the numerous types of architecture being explored for quantum computers are systems utilizing ion traps, in which quantum bits (qubits) are formed from the electronic states of trapped ions and coupled through the Coulomb interaction. Although the elementary requirements for quantum computation have been demonstrated in this system, there exist theoretical and technical obstacles to scaling up the approach to large numbers of qubits. Therefore, recent efforts have been concentrated on using quantum communication to link a number of small ion-trap quantum systems. Developing the array-based approach, we show how to achieve massively parallel gate operation in a large-scale quantum computer, based on techniques already demonstrated for manipulating small quantum registers. The use of decoherence-free subspaces significantly reduces decoherence during ion transport, and removes the requirement of clock synchronization between the interaction regions.

A quantum computer is a device that prepares and manipulates quantum states in a controlled way, offering significant advantages over classical computers in tasks such as factoring large numbers¹ and searching large databases². The power of quantum computing derives from its scaling properties: as the size of these problems grows, the resources required to solve them grow in a manageable way. Hence a useful quantum computing technology must allow control of large quantum systems, composed of thousands or millions of qubits.

The first proposal for ion-trap quantum computation involved confining a string of ions in a single trap, using their electronic states as qubit logic levels, and transferring quantum information between ions through their mutual Coulomb interaction³. All the elementary requirements for quantum computation⁴—including efficient quantum state preparation^{5–7}, manipulation^{7–10} and read-out^{7,11,12}—have been demonstrated in this system. But manipulating a large number of ions in a single trap presents immense technical difficulties, and scaling arguments suggest that this scheme is limited to computations on tens of ions^{13–15}. One way to escape this limitation involves quantum communication between a number of small ion-trap quantum registers. Recent proposals along these lines that use photon coupling^{16–18} and spin-dependent Coulomb interactions¹⁹ have not yet been tested in the laboratory. The scheme presented here, however, uses only quantum manipulation techniques that have already been individually experimentally demonstrated.

The quantum CCD

To build up a large-scale quantum computer, we have proposed a 'quantum charge-coupled device' (QCCD) architecture consisting of a large number of interconnected ion traps. By changing the operating voltages of these traps, we can confine a few ions in each trap or shuttle ions from trap to trap. In any particular trap, we can manipulate a few ions using the methods already demonstrated, while the connections between traps allow communication between sets of ions¹³. Because both the speed of quantum logic gates²⁰ and the shuttling speed are limited by the trap strength, shuttling ions between memory and interaction regions should consume an acceptably small fraction of a clock cycle.

Figure 1 shows a diagram of the proposed device. Trapped ions storing quantum information are held in the memory region. To perform a logic gate, we move the relevant ions into an interaction region by applying appropriate voltages to the electrode segments. In the interaction region, the ions are held close together, enabling

the Coulomb coupling necessary for entangling gates^{3,21}. Lasers are focused through the interaction region to drive gates. We then move the ions again to prepare for the next operation.

We can realize the trapping and transport potentials needed for the QCCD using a combination of radio-frequency (r.f.) and quasistatic electric fields. Figure 1 shows only the electrodes that support the quasistatic fields. By varying the voltages on these electrodes, we confine the ions in a particular region or transport them along the local trap axis, which lies along the thin arrows in Fig. 1. Two more layers of electrodes lie above and below the static electrodes, as shown in Fig. 2. Applying r.f. voltage to the outer layers creates a quadrupole field that confines the ions transverse to the local trap axis by means of the ponderomotive force²². This geometry allows stable transport of the ions around 'T' and 'X' junctions, so we can build complex, multiply connected trap structures.

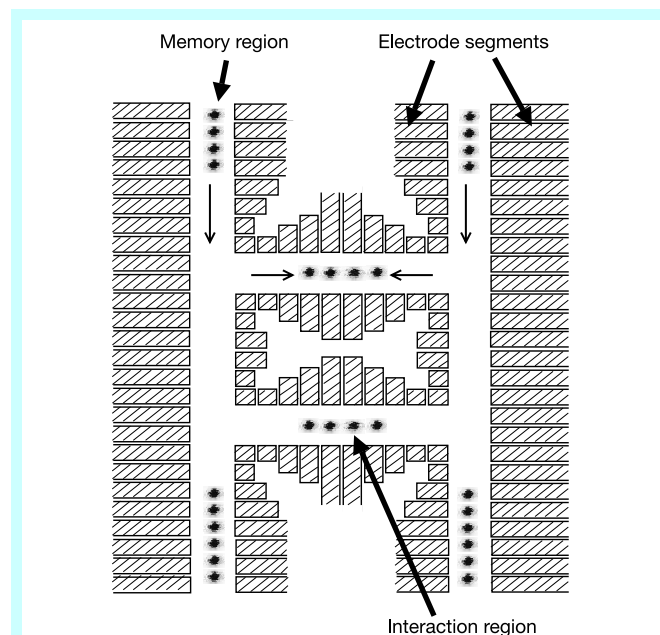


Figure 1 Diagram of the quantum charge-coupled device (QCCD). Ions are stored in the memory region and moved to the interaction region for logic operations. Thin arrows show transport and confinement along the local trap axis.

Electrode structures for the QCCD are relatively easy to fabricate. A number of ion traps have been built by laser-machining slits in alumina wafers and evaporating gold electrodes onto the alumina²³. These traps have geometries similar to that needed for the QCCD, and have spacings between the static and r.f. electrodes of fractions of a millimetre, allowing confinement frequencies up to 20 MHz for r.f. field frequencies of ~ 250 MHz. Similar electrode structures are currently being constructed from heavily doped silicon using standard microfabrication techniques²⁴. Here the silicon acts as the conductive electrode material, while glass spacers anodically bonded to the silicon insulate the r.f. electrodes from the static electrodes.

A first step towards a QCCD has been taken at the National Institute of Standards and Technology (NIST) by constructing a pair of interconnected ion traps; the individual traps are similar to those used in previous work²³ and are separated by 1.2 mm. Efficient coherent transport of a qubit between the two traps was demonstrated by performing a Ramsey-type experiment involving the two traps, where no loss of contrast within the experimental error ($\sim 0.6\%$) was observed²⁵. Transport times were as short as ~ 50 μ s, with corresponding ion velocities greater than 10 m s⁻¹ (see also ref. 26). The transport did not cause any heating of the ion motion or shortening of ion lifetime in the trap.

To maximize the clock speed of the QCCD, we need to transport ions quickly. However, the entangling gate demonstrated in previous work at NIST^{9,21} has low error only for ions cooled near the quantum ground state. To recool the ions after transport and to counteract the effects of heating²³, we propose to use sympathetic cooling of the ions used for quantum logic by another ion species^{13,27,28}. Confining both species in the interaction region lets us use the cooling species as a heat sink, with the Coulomb interaction transferring energy between the two species, as experimentally demonstrated in refs 29–31.

Decoherence

Whereas the decoherence and gate errors in single-trap quantum registers have already been characterized, additional decoherence can occur during ion transport. For instance, the energy splitting of the qubit states of an ion depends on the magnetic field at the ion through the Zeeman effect. During ion transport, the spatial variations of the magnetic field strength along the transport path cause the qubit states to acquire a path-dependent relative phase α , so that, for example, $|\downarrow\rangle + |\uparrow\rangle \rightarrow |\downarrow\rangle + e^{i\alpha}|\uparrow\rangle$ over transport. If we do not know α , we have lost the phase information, effectively dephasing the quantum state. But knowing α for all relevant paths is tantamount to characterizing the magnetic field on micrometre

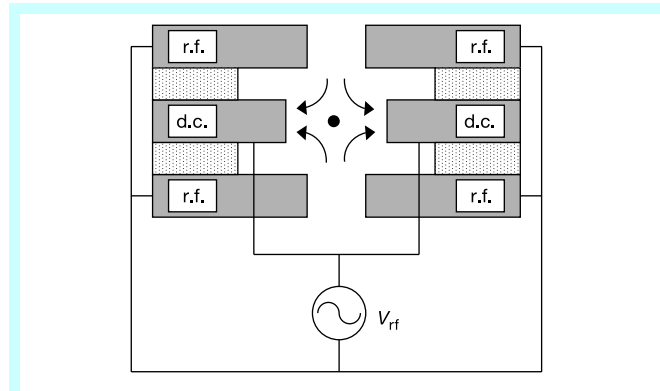


Figure 2 Configuration of radio-frequency (r.f.) and static (d.c.) electrodes for the QCCD. Dotted regions indicate insulating spacers. Applying high r.f. voltage to the two outer electrode layers while keeping the inner layer at r.f. ground creates the r.f. quadrupole field shown by the arrows. This field provides trapping potentials for confinement transverse to the local trap axis, which points perpendicular to the page and is located at the central black dot. View in Fig. 1 is from the top of this figure.

length scales across the entire device, a very difficult task.

Retaining phase information during a computation also requires accurate positioning of the ions in the interaction regions. As the logic gate parameters depend on the phase of the driving laser fields at the ion positions, the ion positions and laser path lengths must be controllable with accuracies much better than an optical wavelength. Although this does not place unreasonable constraints on the accuracy of the voltage sources driving the QCCD electrodes, stray electric fields emanating from the electrodes or mechanical vibration of the QCCD can readily move the ions a fraction of a wavelength from their nominal positions, effectively adding a relative phase α of the type considered above.

Decoherence-free encoding

We can reduce these sources of decoherence by several orders of magnitude by encoding each qubit into a decoherence-free subspace (DFS) of two ions^{32–34}. This DFS is spanned by the states $|0\rangle = |\downarrow\downarrow\rangle$, $|1\rangle = |\uparrow\uparrow\rangle$ of the ions. We refer to the ions as ‘physical qubits’ and call the effective two-level system formed by $|0\rangle$, $|1\rangle$ a ‘logical qubit’. If the state $|\uparrow\rangle$ of ion j acquires a phase α_j over and above the predicted phase α_0 for the transport path, we see

$$|0\rangle + |1\rangle \rightarrow e^{i\alpha_1}|\downarrow\downarrow\rangle + e^{i\alpha_2}|\uparrow\uparrow\rangle = |0\rangle + e^{i\Delta\alpha}|1\rangle \quad (1)$$

where we write $\Delta\alpha \equiv \alpha_2 - \alpha_1$. If $\Delta\alpha = 0$, we see that the DFS logical qubit is unaffected by the acquired phase. Here the phases α_j are themselves unknown, but the ions acquire the same unknown phase, a process called collective dephasing. The logical qubit decoheres only insofar as the dephasing fails to be collective.

As an example, assume that the physical qubit energies have a linear dependence on magnetic field and that the field varies linearly over an extended QCCD device of size 10 cm. If each pair of physical qubits comprising a logical qubit is separated on average by 10 μ m, a logical qubit dephases more slowly than a physical qubit by a factor of 10^4 . Again, Stark shifts of the qubit levels can be induced by the electric fields that push the ions from place to place, but their dephasing effects on the logical qubits are also suppressed. In general, the effect of any external field varying over a length scale L_{ext} and inducing energy shifts of p th power in the field is suppressed by a factor $(L_{\text{ext}}/d)^p$ for DFS encoding. A DFS-encoded qubit is therefore robust against decoherence incurred during transport.

We can also perform universal quantum logic in the DFS, as we now show. If we hold two ions in an interaction region, we can use the entangling gate of refs 9 and 21 to apply the operator

$$U_2(\theta, \phi_1, \phi_2) = \cos\theta[I_1 \otimes I_2] + i \sin\theta[X_{\phi_1} \otimes X_{\phi_2}] \\ = \cos\theta I^{\text{DFS}} + i \sin\theta X_{\Delta\phi_{12}}^{\text{DFS}} \quad (2)$$

$$X_{\phi} \equiv X \cos \phi + Y \sin \phi \quad (3)$$

where X , Y are the Pauli operators and the superscript ‘DFS’ indicates that the operator acts in the DFS logical basis. Though the individual phases ϕ_1 , ϕ_2 depend sensitively on the path-length differences of the driving lasers over the macroscopic distance from the lasers to the ions¹³, the DFS gate phase $\Delta\phi_{12} = \phi_1 - \phi_2$ depends only on the microscopic path length of the driving laser between the two ions, which can be readily controlled by small adjustments of the trap voltages^{9,12,24}. The DFS encoding makes the computation insensitive to spatial phase fluctuations; as we will see, it protects against temporal phase fluctuations as well. We can set θ over a wide range of values^{21,24}, so the two-ion entangling gate lets us perform arbitrary rotations of a single logical qubit. Using the same entangling gate on four ions, we can obtain the operator

$$U_4 = \frac{1}{\sqrt{2}}[I_1 \otimes I_2 \otimes I_3 \otimes I_4 - iX_{\phi_1} \otimes X_{\phi_2} \otimes X_{\phi_3} \otimes X_{\phi_4}] \quad (4)$$

$$= \frac{1}{\sqrt{2}}[I^{\text{DFS}} \otimes I^{\text{DFS}} - iX_{\Delta\phi_{12}}^{\text{DFS}} \otimes X_{\Delta\phi_{34}}^{\text{DFS}}] \quad (5)$$

where ions 1 and 2 encode one logical qubit and ions 3 and 4 encode another, and we write $\Delta\phi_{34} = \phi_3 - \phi_4$. This operator is equivalent to an XOR in the logical basis up to rotations of single logical qubits³⁵, so the operators of equations (2) and (5) suffice for universal quantum logic.

To use the DFS encoding in a large-scale quantum computation, we initialize the ions in pairs to the state $|\downarrow\uparrow\rangle$. Each pair of ions remains in the DFS through the quantum computation, so the logical qubits resist transport decoherence and all other types of collective dephasing. Read-out of the DFS qubit is straightforward, as we need only discriminate between $|\downarrow\uparrow\rangle$ and $|\uparrow\downarrow\rangle$. All the operators needed for universal quantum logic in the DFS have already been experimentally implemented^{9,24}, so we should be able to use the DFS encoding in a large-scale quantum computer. Notably, all logic gate operations can be accomplished by uniformly illuminating the ions in the interaction region, removing the need for tightly focused laser beams.

Logic gate synchronization

The DFS encoding also removes the requirement of clock synchronization between logic gates, a major but little-recognized obstacle to large-scale parallel quantum computation. As the energy levels of our physical qubits are non-degenerate, we must keep track of the resulting phase accumulation to preserve the quantum information in the physical qubit basis³⁴. Parallel operations taking place in many interaction regions thus require clocks that remain synchronized over the whole computation time³⁶. Synchronization can become very difficult for many qubits: for a transition frequency ω_0 between $|\downarrow\rangle$ and $|\uparrow\rangle$, the two components of the state $|\downarrow\rangle^N + |\uparrow\rangle^N$ acquire a significant relative phase in a time $\sim 1/(N\omega_0)$. To maintain phase stability of the computation, we therefore require a frequency reference with fractional frequency stability much better than $\lesssim 1/(N\omega_0\tau)$ at an averaging time τ equal to the duration of the quantum computation.

To be concrete, we consider trapped $^{40}\text{Ca}^+$ ions as qubits, with the ground $S_{1/2}$ state and metastable $D_{5/2}$ states as logic levels. This system is being investigated for quantum computation by a number of groups^{7,15,37}. Here the transition frequency is 412 THz, comparable to the 533 THz operating frequency of the currently most stable laser oscillator³⁸, which has a fractional frequency instability of 3×10^{-16} as 1 s averaging time. If the computation takes about 1 s, equal to the lifetime of the metastable $D_{5/2}$ state, we see that current technology barely provides the appropriate phase stability for even one ion. Of course, this argument can be regarded as too simplistic because the requirements on phase stability can be reduced by invoking error correction³⁹; however, this comes at the cost of increased overhead¹⁸. For qubit levels with a transition frequency in the microwave regime, local oscillators of the required phase stability exist, but the optical path lengths between the driving lasers and each of the interaction regions must be stable at the nanometre level over the course of the computation¹³, a daunting task for a 10 cm QCCD device.

On the other hand, as the logic levels of a DFS-encoded qubit are degenerate, we do not need phase synchronization at all to perform a logic operation within the DFS. Operations in the DFS are also independent of the optical path lengths of the driving lasers, because the phases $\Delta\phi_{12}$, $\Delta\phi_{34}$ in equations (2) and (5) depend only on the distance between the two ions comprising a logical qubit. The universal gate-set constructed above allows us to perform highly parallel computations in the DFS without synchronization between gates separated in time or space. Similar considerations would apply to other quantum computing architectures. $\square$

doi:10.1038/nature00784.

- Shor, P. W. In *Proc. 35th Annu. Symp. on the Foundations of Computer Science* (ed. Goldwasser, S.) 124–134 (IEEE Computer Society, Los Alamitos, 1994).
- Grover, L. K. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79**,

325–328 (1997).

- Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
- DiVincenzo, D. P. In *Scalable Quantum Computers* (eds Braunstein, S. L. & Lo, H. K.) 1–14 (Wiley-VCH, Berlin, 2001).
- Monroe, C. *et al.* Resolved-sideband Raman cooling of a bound atom to the 3D zero-point energy. *Phys. Rev. Lett.* **75**, 4011–4014 (1995).
- King, B. E. *et al.* Cooling the collective motion of trapped ions to initialize a quantum register. *Phys. Rev. Lett.* **81**, 1525–1528 (1998).
- Roos, C. *et al.* Quantum state engineering on an optical transition and decoherence in a Paul trap. *Phys. Rev. Lett.* **83**, 4713–4716 (1999).
- Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
- Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
- Nägerl, H. C. *et al.* Laser addressing of individual ions in a linear ion trap. *Phys. Rev. A* **60**, 145–148 (1999).
- Blatt, R. & Zoller, P. Quantum jumps in atomic systems. *Eur. J. Phys.* **9**, 250–256 (1988).
- Rowe, M. A. *et al.* Experimental violation of a Bell's inequality with efficient detection. *Nature* **409**, 791–794 (2001).
- Wineland, D. J. *et al.* Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. NIST* **103**, 259–328 (1998).
- Hughes, R. J., James, D. F. V., Knill, E. H., Laflamme, R. & Pechesek, A. G. Decoherence bounds on quantum computation with trapped ions. *Phys. Rev. Lett.* **77**, 3240–3243 (1996).
- Enzer, D. G. *et al.* In *Experimental Implementation of Quantum Computation '01* (ed. Clark, R.) (Rinton, Princeton, 2001).
- Pellizzari, T., Gardiner, S. A., Cirac, J. I. & Zoller, P. Decoherence, continuous observation, and quantum computing: A cavity QED model. *Phys. Rev. Lett.* **75**, 3788–3791 (1995).
- DeVoe, R. G. Elliptical ion traps and trap arrays for quantum computation. *Phys. Rev. A* **58**, 910–914 (1998).
- Steane, A. M. & Lucas, D. M. Quantum computing with trapped ions, atoms and light. *Fortsch. Phys.* **48**, 839–858 (2000).
- Cirac, J. I. & Zoller, P. A scalable quantum computer with ions in an array of microtraps. *Nature* **404**, 579–581 (2000).
- Steane, A. *et al.* Speed of ion-trap quantum-information processors. *Phys. Rev. A* **62**, 042305 (2000).
- Sørensen, A. & Mølmer, K. Entanglement and quantum computation with ions in thermal motion. *Phys. Rev. A* **62**, 022311 (2000).
- Paul, W. Electromagnetic traps for charged and neutral particles. *Rev. Mod. Phys.* **62**, 531–540 (1990).
- Turchette, Q. A. *et al.* Heating of trapped ions from the quantum ground state. *Phys. Rev. A* **61**, 063418 (2000).
- Kielinski, D. *Entanglement and Decoherence in a Trapped-ion Quantum Register*. Thesis, Univ. Colorado (2001); available at (<http://www.boulder.nist.gov/timefreq/ion/qucomp/papers.htm>).
- Rowe, M. A. & Zoller, P. Transport of quantum states and separation of ions in a dual RF ion trap. Preprint quant-ph/0205094 at (<http://xxx.lanl.gov>) (2002).
- Guthöhrlein, G. R., Keller, M., Hayasaka, K., Lange, W. & Walther, H. A single ion as a nanoscopic probe of an optical field. *Nature* **414**, 49–51 (2001).
- Kielinski, D. *et al.* Sympathetic cooling of trapped ions for quantum logic. *Phys. Rev. A* **61**, 032310 (2000).
- Morigi, G. & Walther, H. Two-species Coulomb chains for quantum information. *Eur. Phys. J. D* **13**, 261–269 (2001).
- Larson, D. J., Bergquist, J. C., Bollinger, J. J., Itano, W. M. & Wineland, D. J. Sympathetic cooling of trapped ions: A laser-cooled two-species nonneutral ion plasma. *Phys. Rev. Lett.* **57**, 70–73 (1986).
- Rohde, H. *et al.* Sympathetic ground-state cooling and coherent manipulation with two-ion crystals. *J. Opt. B* **3**, S34–S41 (2001).
- Blinov, B. B. *et al.* Sympathetic cooling of trapped Cd^+ isotopes. Preprint quant-ph/0112084 at (<http://xxx.lanl.gov>) (2001).
- Lidar, D. A., Chuang, I. L. & Whaley, K. B. Decoherence-free subspaces for quantum computation. *Phys. Rev. Lett.* **81**, 2594–2597 (1998).
- Duan, L. M. & Guo, G. C. Reducing decoherence in quantum-computer memory with all quantum bits coupling to the same environment. *Phys. Rev. A* **57**, 737–741 (1998).
- Kielinski, D. *et al.* A decoherence-free quantum memory using trapped ions. *Science* **291**, 1013–1015 (2001).
- Sørensen, A. & Mølmer, K. Quantum computation with ions in thermal motion. *Phys. Rev. Lett.* **82**, 1971–1974 (1999).
- van Enk, S. J. The physical meaning of phase and its importance for quantum teleportation. *J. Mod. Opt.* **48**, 2049–2054 (2001).
- Steane, A. The ion-trap quantum information processor. *Appl. Phys. B* **64**, 623–643 (1997).
- Young, B. C., Cruz, F. C., Itano, W. M. & Bergquist, J. C. Visible lasers with subhertz linewidths. *Phys. Rev. Lett.* **82**, 3799–3802 (1999).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* 425–493 (Cambridge Univ. Press, Cambridge, 2000).

Acknowledgements

We acknowledge the experimental contributions of the NIST Ion Storage group, and also J. Beall for assistance with microfabrication. We thank D. Leibfried and M. A. Rowe for comments on the manuscript. D.K. and D.J.W. were supported by the US National Security Agency (NSA), Advanced Research and Development Activity (ARDA) and the Office of Naval Research. C.M. was supported by the US NSA, ARDA and the National Science Foundation ITR programme.

Correspondence and requests for materials should be addressed to D.K. (e-mail: utonium@mit.edu).

Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution

Dmitry G. Vassylyev^{*†}, Shun-ichi Sekine^{*†}, Oleg Laptenko[‡], Jookyung Lee[‡], Marina N. Vassylyeva^{†§}, Sergei Borukhov[‡] & Shigeyuki Yokoyama^{*†§||}

^{*} Cellular Signaling Laboratory, and [†] Structurome Research Group, RIKEN Harima Institute at Spring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan

[‡] Department of Microbiology, SUNY Health Science Center, 450 Clarkson Avenue, Brooklyn, New York 11203, USA

[§] RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

^{||} Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

In bacteria, the binding of a single protein, the initiation factor σ , to a multi-subunit RNA polymerase core enzyme results in the formation of a holoenzyme, the active form of RNA polymerase essential for transcription initiation. Here we report the crystal structure of a bacterial RNA polymerase holoenzyme from *Thermus thermophilus* at 2.6 Å resolution. In the structure, two amino-terminal domains of the σ subunit form a V-shaped structure near the opening of the upstream DNA-binding channel of the active site cleft. The carboxy-terminal domain of σ is near the outlet of the RNA-exit channel, about 57 Å from the N-terminal domains. The extended linker domain forms a hairpin protruding into the active site cleft, then stretching through the RNA-exit channel to connect the N- and C-terminal domains. The holoenzyme structure provides insight into the structural organization of transcription intermediate complexes and into the mechanism of transcription initiation.

The DNA-dependent RNA polymerase (RNAP) is the principal enzyme of the transcription process, and is a final target in many regulatory pathways that control gene expression in all living organisms. In bacteria, RNAP is responsible for the synthesis of all RNAs in the cell (messenger RNA, ribosomal RNA, transfer RNA, and so on). The bacterial RNAP exists in two forms: core and holoenzyme. The core enzyme has a relative molecular mass of about 400,000, and consists of five subunits: α -dimer (α_2), β , β' and ω . They are evolutionarily conserved in sequence, structure and function from bacteria to humans^{1,2}.

The transcription cycle in bacterial cells can be divided into three main stages: initiation, elongation and termination. Although it is catalytically active, the core enzyme is incapable of initiating transcription efficiently and with specificity. For this, it must bind an initiation factor, σ , to form a holoenzyme that can recognize specific DNA sequences (promoters)^{3,4}. During initiation, the holoenzyme specifically binds to two conserved hexamers in the promoter at nucleotide (nt) positions -35 and -10 relative to the transcription start site ($+1$), to form a closed promoter complex. Then, it unwinds the double-stranded DNA (dsDNA) around the -10 region (between nt -12 and $+2$), resulting in the open promoter complex, and initiates transcription in the presence of nucleoside triphosphate substrates⁵. After the synthesis of an RNA 9–12 nt long, of which 8 or 9 are base-paired with the DNA template strand (DNA–RNA hybrid), the transcription complex passes from the initiation to elongation stage⁶. This transition is characterized by the escape of RNAP from the promoter, the dissociation of σ from the core, and the formation of a highly processive ternary elongation complex. An alternative mechanism has been proposed, in which σ factor remains bound to the core in the ternary complex during the elongation stage^{7,8}. During the transition from initiation to elongation, RNAP undergoes a significant conformational change^{5,9}. The σ factor has a central role in initiation, being directly involved in promoter recognition, DNA melting and promoter escape and clearance^{4,10}.

Among several distinct σ factors^{4,11}, $\sigma 70$ is primarily responsible for the transcription of most ‘housekeeping’ genes in the cell during

exponential growth. The family of related $\sigma 70$ proteins shares four regions of sequence homology, designated 1–4, which are further divided into several subregions¹². The regions 4.2, 2.3–2.4 and 2.5 were shown to recognize the -35 , -10 , and the so-called ‘extended -10 ’ (around nt -17 to -13) elements of the promoter, respectively^{13–15}. In addition, regions 2.3–2.4 are known to interact specifically with the -10 region of the nontemplate strand in the open promoter complex^{16,17}. Finally, genetic and biochemical data have shown that the portions of σ that contact the core include most of these conserved regions^{18–20}.

The crystal structures of the *Thermus aquaticus* RNAP core²¹, the yeast *Saccharomyces cerevisiae* RNAP (polymerase II, Pol II)²², and the Pol II ternary elongation complex with a 9-base pair (bp) DNA–RNA hybrid²³ revealed the high degree of structural similarity between the prokaryotic and eukaryotic RNAPs. Together with information gained from a wide range of biochemical, biophysical and genetic studies, these structures allow a plausible modelling of bacterial elongation and open promoter transcription complexes^{24,25}.

However, many aspects of the σ -dependent transcription initiation still remain obscure. These include: the promoter recognition by the holoenzyme; the formation of a closed promoter complex and its transition to an open complex; and the promoter escape and dissociation of σ factor from RNAP. The high-resolution structure of the *T. thermophilus* RNAP holoenzyme reported here provides information crucial to addressing these questions. It allows us to identify the structural changes in the core induced by σ binding, and to correlate them with the known functional, biologically relevant properties of the holoenzyme. Finally, the holoenzyme structure substantially improves the atomic model of the RNAP core, which is currently available at a modest resolution.

Structure determination and overall structure

The holoenzyme crystals belong to the $P3_2$ space group with unit cell dimensions $a = b = 236.35$, $c = 249.04$ Å. The structure was determined by molecular replacement and refined against data to 2.6 Å resolution (Fig. 1a, b; see also Supplementary Information).

The holoenzyme maintains the overall crab-claw-like shape (Fig. 1c) that was observed in the structures of the *T. aquaticus* core and *S. cerevisiae* Pol II^{21,22}. However, the presence of the σ 70 subunit and the large portion of the nonconserved N-terminal domain of the β' subunit in the model greatly extends the β' pincer, thus making the analogy to a crab claw even more fitting. The σ 70 subunit is located almost entirely on the core surface, except for a short segment (σ 313–342), which is buried within the core molecule. The binding of σ to the RNAP core does not seem to result in the widening of the DNA-binding channel, as was previously thought. On the contrary, certain structural elements of σ reduce the available space of functionally important cavities in RNAP that are presumed to accommodate the transcription bubble and the RNA product. These elements include: (1) the second

N-terminal domain of σ , which bridges the β and β' pincers of the claw on the upstream DNA side of RNAP, creating a wall that blocks the opening of the DNA-binding channel; (2) a hairpin-like motif of σ , which protrudes into the cleft of the active site, towards the possible location of the RNA–DNA hybrid; and (3) an extended fragment following the hairpin-like motif, which occupies the RNA-exit channel.

The core-to-holoenzyme conversion is accompanied by changes in all structural domains of the core enzyme. The most noticeable include the β G flap domain (β 702–828), the α I subunit, the large N-terminal portion of β (β 20–395), and the β' coiled-coil domain (β' 540–585) (Fig. 2). The RNA-exit channel is most significantly affected by σ 70 binding. The channel is constricted by the β flap domain, which is shifted by about 5–6 Å towards the σ C-terminal domain compared with the *T. aquaticus* core²¹. The change in the orientation of the ‘flap tip’ helix, which is now squeezed between three α -helices of σ (Fig. 2), is even more dramatic (~11 Å shift). The change in the orientation of individual domains of the RNAP subunits (α I and β 20–395) that are adjacent to the β flap domain seems to mirror the flap domain rotation.

In the crystal structure, we observed many Mg^{2+} ions bound to holoenzyme, most of which formed a coat on the protein surface. These metal ions could be important in the binding and bending of the DNA molecule, which is proposed to be wrapped around RNAP during transcription²⁶.

Structure of σ 70

The atomic model of the σ 70 subunit lacks the disordered N-terminal domain (σ 1–73), which includes the conserved σ region 1.1. This is a self-inhibitory domain, which is known to mask the DNA-binding regions of σ before it binds the core²⁷. Its precise role in transcription initiation is not understood.

The modelled structure of σ 70 consists entirely of α -helices connected by either turns or loops (Fig. 3). It can be divided into four structural domains: N-terminal domain 1 (ND1), N-terminal domain 2 (ND2), ‘linker’ domain (LD) and C-terminal domain (CD).

ND1 consists of eight α -helices (σ 74–254) comprising four helix–turn–helix (HtH) motifs: HtH1 (σ 75–121), HtH2 (σ 94–137), HtH3 (σ 123–147) and HtH4 (σ 152–203) (Figs 3 and 4a). This domain encompasses region 1.2 up to the N-terminal half of region 2.4, including the nonconserved segment between regions 1 and 2. ND1 has a U-shaped structure, and is connected to ND2 (σ 261–312) by a short linker loop (σ 255–260), which lies at the C terminus of region 2.4.

ND2, corresponding to regions 2.4–2.5 and 3.1, consists of three α -helices that fold into an α -helical bundle (Figs 3 and 4a). The C-terminal helix of ND1 and the N-terminal helix of ND2 (σ 234–281) form a V-shaped structure (Fig. 4a) near the opening of the upstream DNA-binding channel. The inner surface of the V is lined by the residues from the C- and N-terminal α -helices of ND1 and ND2, respectively. These α -helices correspond to regions 2.3–2.5 (Fig. 3c), which are known to be involved in the specific binding of σ to the –10 and extended –10 elements of the promoter^{14,15}.

The 30-residue LD (σ 313–339; region 3.2) intervenes between the globular N- and C-terminal portions of σ , and has mostly an extended, unfolded conformation (Figs 3b and 4b). Roughly at its midpoint, the LD forms a hairpin loop (σ 318–329) that protrudes into the active site cleft, between the ‘rudder’ loop (β' 584–601) and the ‘lid’ loop (β' 526–538), which are believed to interact directly with nucleic acids^{21,22}. The rest of the LD stretches along the RNA-exit channel towards σ 's CD.

The CD (σ 340–423), which includes conserved regions 4.1 and 4.2, contains four α -helices, which are arranged as a pair of HtH motifs (HtH5, σ 341–376; and HtH6, σ 387–418) (Figs 3 and 4c). The conserved σ region 4 was shown to recognize the –35 promoter element specifically^{28,29}. The CD is located on the core surface in the

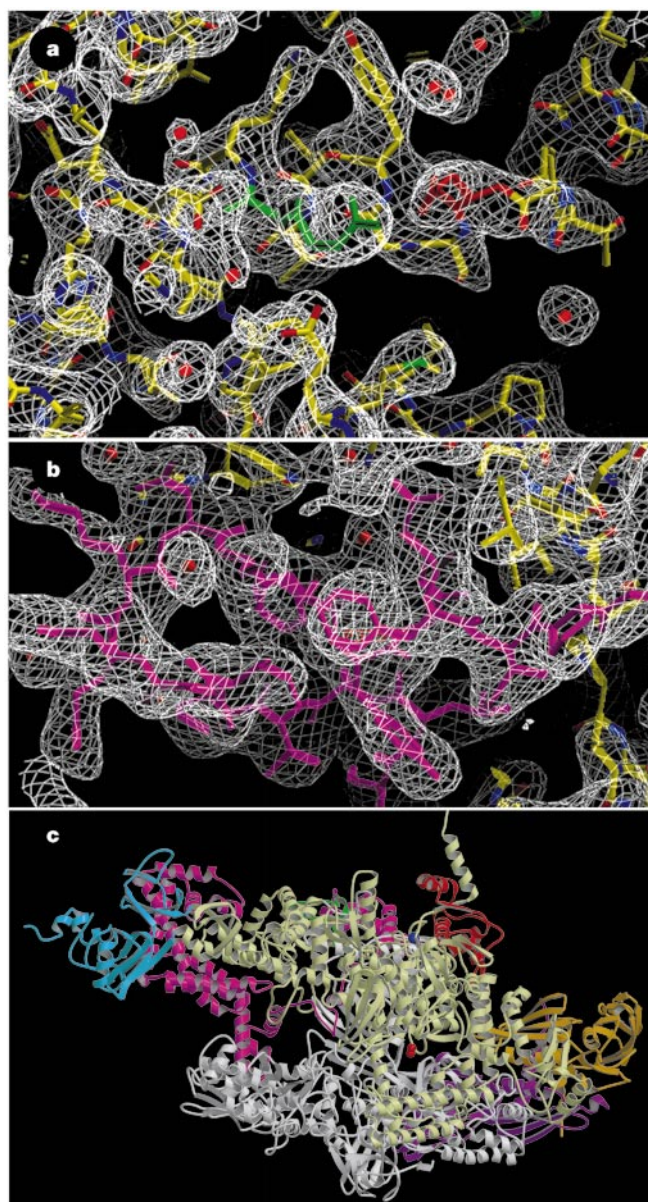


Figure 1 Holoenzyme crystal structure. **a**, **b**, The β' F bridge helix (**a**) and the hairpin loop of σ (**b**). The $|2F_o - F_c|$ electron density map (1.3 σ contour level; white) is superimposed on the atomic model. Protein atoms have atom-dependent colours, except for the σ -subunit (magenta) and the β' residues Asp 1,090 (red) and Arg 1,096 (green). Water molecules are represented by red spheres. **c**, The overall structure. The colours of the subunits are: β , khaki; β' , white (β' 163–452, cyan; β' Zn finger, green); α I, blue; α II, light orange; σ , magenta; and ω , red. Two catalytic Mg^{2+} (red) and two Pb^{2+} (Zn^{2+}) ions (blue) are shown as spheres.

vicinity of the opening of the RNA-exit channel. This site is about 57 Å from the N-terminal half of the protein, containing regions 2.3–2.5.

All of the conserved sequence motifs between the *T. thermophilus* σ 70 and the *Escherichia coli* σ 70 fragment³⁰ are also conserved in three dimensions with a root-mean-square deviation (r.m.s.d.) of 2.9 Å over 125 α -carbon (C α) atoms³¹. Therefore, it is likely that the *E. coli* σ 70 conformation and mode of core binding are similar to those of the *T. thermophilus* protein. Interestingly, we detected an intrinsic structural similarity between the CD and the nonconserved portion of ND1 in *T. thermophilus* σ (Fig. 4d), although they do not exhibit any sequence homology. Structural similarity also exists, to a lesser extent, between the CD and the nonconserved region of the *E. coli* σ 70 fragment (σ 274–338). This observation raises the possibility that these structural segments have evolved from a common ancestor. Their structural similarity and the presence of HtH motifs suggest that the nonconserved ND1 region of σ might have served as an alternative DNA-binding site at some stage of evolution.

β' zinc finger and N-terminal domains

The previously reported structure of the *T. aquaticus* core enzyme lacks the 'zinc finger' domain (β' 53–81) and the N-terminal non-conserved region of the β' subunit (β' 163–452). These domains are fixed in the holoenzyme structure.

The zinc finger domain has no secondary structure, and folds into a U-shaped structure, with two long loops harbouring one Pb²⁺ ion (Zn²⁺ in the native enzyme) between them. In the holoenzyme, this domain is located between the N- and C-terminal domains of σ 70, in the vicinity of σ 's CD, which implies a potential role in promoter binding (Fig. 4a, c).

We traced a large portion of the β' N-terminal nonconserved domain, spanning residues β' 163–251 and β' 364–452. This domain consists mostly of β -strands, with one α -helix near its N terminus, and two short α -helices at the end of its N-terminal portion, just before the gap. The nonconserved β' domain is close to the σ ND1 domain, which suggests that its primary function is as an anchor for the σ 70 subunit (Figs 1c and 4a).

Interactions between σ 70 and the core

In summary, among the four structural domains of σ 70, ND1 makes the most extensive contacts with the RNAP core. Its U-shaped structure folds around the nearby α -helical coiled-coil of β' (β' 540–585)^{18,20,32,33}. Residues from both nonconserved and conserved (1.2, 2.1, 2.2, 2.4) segments of ND1 make multiple polar interactions with these β' α -helices (Fig. 4a). The outer surface of the U interacts with the N-terminal portion of the β' subunit (β' 95–460), which includes a large portion of the nonconserved segment.

The ND2 of σ 70 (Fig. 4a) is loosely bound to the core. Although it bridges the $\beta\beta'$ pincers of the claw, it has only a few direct interactions with the core molecule. However, the binding is enhanced by multiple water-mediated interactions between ND2 and the core. The putative flexibility of ND2 may be important in the –10 hexamer recognition and/or in the discrimination between promoters with –10 or extended –10 elements^{14,15}.

The σ 70 LD makes few contacts at its N terminus with the β' loops that form the 'rudder' and 'lid' structures (Fig. 4b). The hairpin loop (σ 318–329) interacts with the 'lid' loop, as well as the short α -helix (β' 612–619) that follows the rudder in β' . The C-terminal half of the LD, which is entirely buried within the RNA-exit channel, makes a couple of contacts with the N-terminal segment of β' and the C terminus of β .

The interaction of σ CD with the core is limited to its contacts with the flap tip helix of the β subunit (β 768–779) on the one side, and the β' zinc finger domain on the other (Fig. 4c).

Altogether, there are surprisingly few direct polar interactions between the residues of the σ and core molecules, despite their large interface area.

Catalytic site

As in the structures of the *T. aquaticus* core²¹ and *S. cerevisiae* Pol II²², the active site of *T. thermophilus* RNAP is defined by the three invariant aspartate residues of the β' subunit (Asp 739, Asp 741 and Asp 743), located within a highly conserved sequence motif (NADFQGD)³⁴. In the holoenzyme structure, one of the two molecules in the asymmetric unit has one Mg²⁺ ion chelated by three active-site aspartates. The second molecule of the holo-

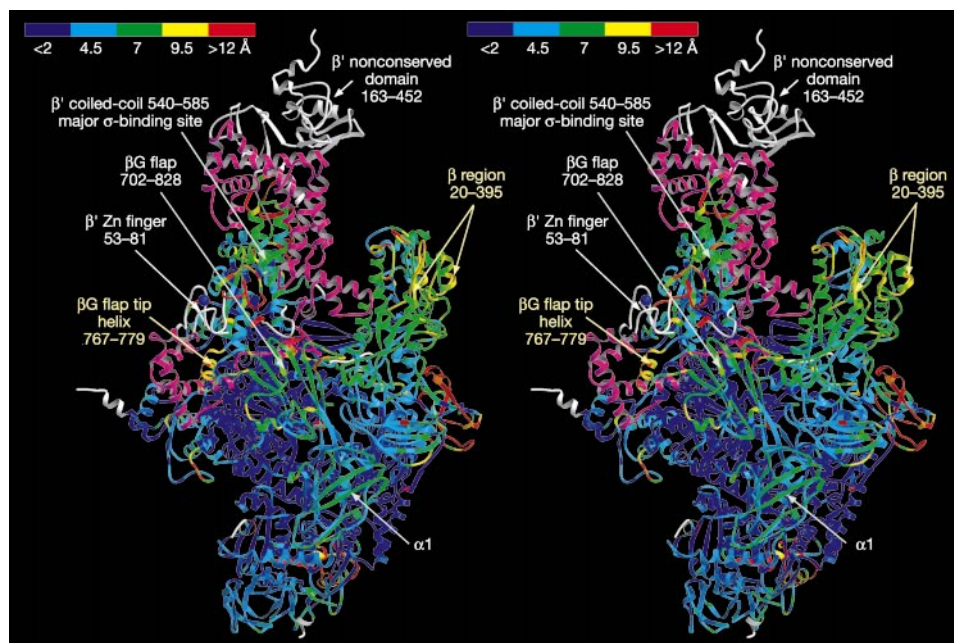


Figure 2 Comparison between the holoenzyme and *T. aquaticus* core structures. In the stereo view shown, the holoenzyme is coloured from blue (<2 Å) to red (>12 Å), according to the deviation in the positions of the C α atoms of the corresponding holoenzyme and core residues. The σ subunit and other structural elements that are

absent from the *T. aquaticus* structure are shown in magenta and white, respectively. Most of the deviations that are more than 12 Å (red) correspond to the mis-assigned regions in the *T. aquaticus* structure. The structures were superimposed by the β' subunits (r.m.s.d. = 1.65 Å over 700 C α atoms).

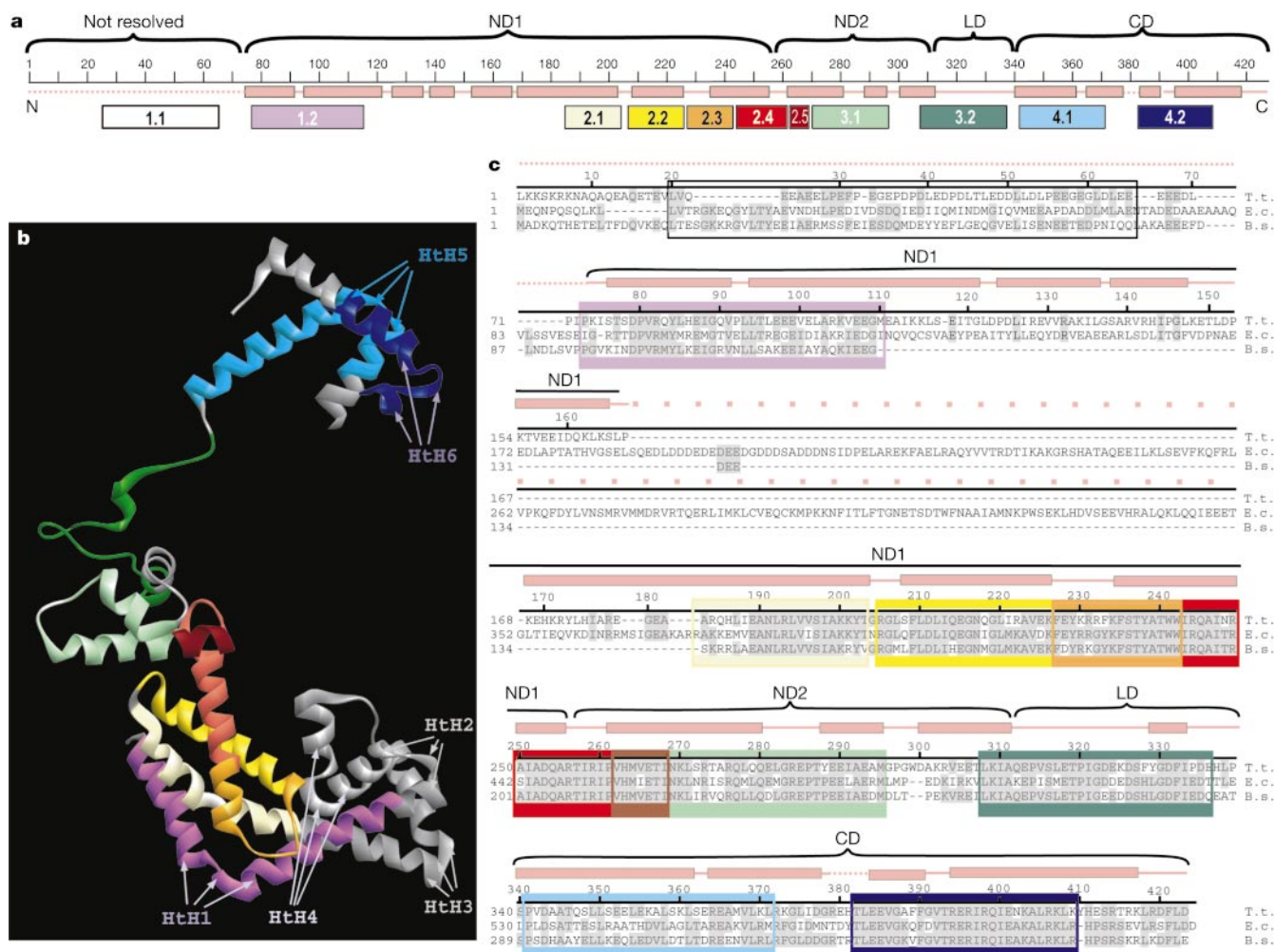


Figure 3 Organization and structure of the σ . **a**, Scheme of structural domains and conserved regions. **b**, Ribbon diagram of σ . The colour coding is the same as in **a** except for the nonconserved regions (grey). **c**, Secondary structure and sequence alignment. Numbers of *T. thermophilus* residues are used. In **a** and **c**, the secondary structure is

indicated in pink as boxes (α -helices), solid lines (coils) and dotted lines (disordered structures). The σ factors from *T. thermophilus* (T.t.), *E. coli* (E.c.) and *Bacillus subtilis* (B.s.) were aligned using the LaserGene program. Identical residues are greyed. Conserved regions of σ (1.1–4.2 in **a**) are boxed and coloured as in **a**.

enzyme contains two Mg^{2+} ions (with B -factors of $26 \text{ \AA}^2$ and $27 \text{ \AA}^2$) bound at the catalytic centre. The same three conserved aspartates, together with several neighbouring water molecules, form a network of interactions responsible for holding these two Mg^{2+} ions (Fig. 5a). Two metal ions are separated by $3.0 \text{ \AA}$, and the distance between each metal ion and the oxygen ligand is in the range of 2.5 – $4.0 \text{ \AA}$. These results are consistent with the two-metal-ion mechanism proposed for all nucleic acid polymerases³⁵.

The bridge helix

The β' bridge helix (β' 1,067–1,102), which separates the internal space occupied by the transcription bubble and the downstream DNA, had somewhat different conformations in the structures of the *T. aquaticus* core and the eukaryotic Pol II^{21–23}. This difference may reflect two possible states of the bridge helix in both enzymes. It has been proposed that the transition between the two conformations triggers the translocation of RNAP during RNA polymerization²². Here, we present a more detailed view of this process, based on the high-resolution structure of the holoenzyme.

The overall shape of the bridge helices in the bacterial holoenzyme and Pol II is nearly identical (r.m.s.d. = $1.1 \text{ \AA}$ over 33 C α atoms), with both helices having a bend of about 28° (Fig. 5b). The conformational difference between the two enzymes is localized to the central region of the helix. In Pol II, the bridge helix maintains a

uniform α -helical conformation, whereas in the holoenzyme, the backbones of two residues (Asp 1,090, Ser 1,091) in the middle are flipped out of the α -helix (not partially unwound, as observed in the *T. aquaticus* core structure). This pair of flipped-out residues could serve as a molecular 'switch' in RNAP translocation. In the holoenzyme structure, the observed unstable, flipped-out conformation of the bridge helix centre is stabilized by intrahelical hydrogen bonding between Asp 1,090 and the invariant Arg 1,096 (Fig. 5c). In the Pol II elongation complex structure (Protein Data Bank (PDB) accessions number 1I6H), the counterpart of this arginine bridges two adjacent phosphates of the template strand DNA at positions i and $i+2$ (conventionally, i and $i+1$ refer to the product-binding and substrate-binding sites, respectively) (Fig. 5d). In addition to the bridge helix, a conserved β' G loop comprising residues 1,238–1,250 may also be involved in the translocation process. This loop is traced in the holoenzyme structure, but is disordered in both the *T. aquaticus* core and Pol II structures^{21,22}. The β' G loop forms part of the floor of the downstream DNA-binding cavity, and is located close to the distorted, central region of the bridge helix (Fig. 5c). If we model the bridge helix in a uniform α -helical conformation (as observed in the Pol II structure), then the Asp 1,090 side chain would clash with a patch of highly conserved hydrophobic residues (Met 1,238, Phe 1,241 and Leu 1,256) located in this flexible β' G loop (Fig. 5d). Thus, it appears that the bridge

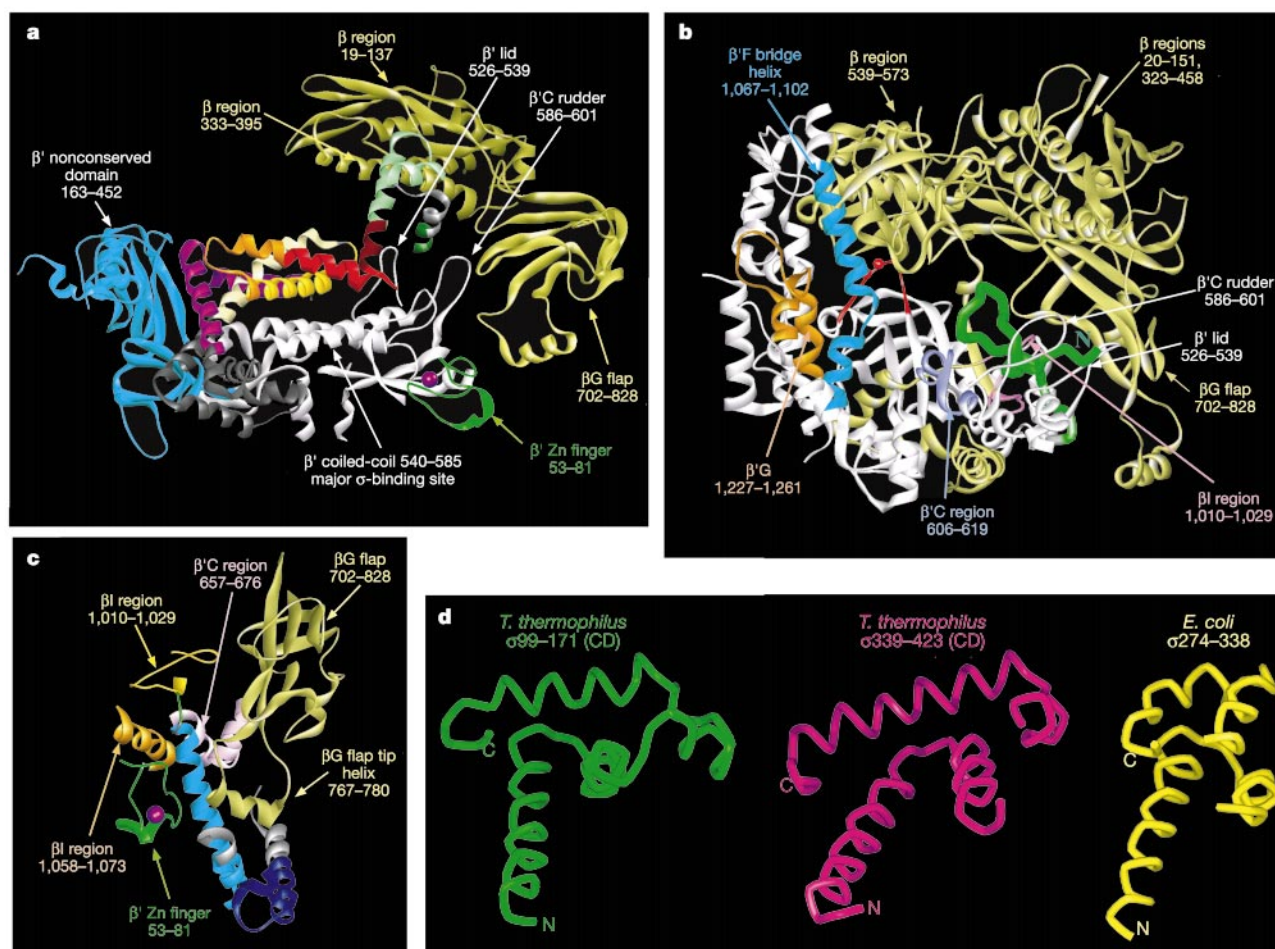


Figure 4 Individual σ domains and their interactions with the core. The conserved regions of σ have the same colours as in Fig. 3a. **a**, The σ domains ND1 and ND2 and the neighbouring core regions are shown. **b**, Partial view of the active centre cleft, showing the buried location of σ LD (green) and its hairpin loop. The catalytic loop and the Mg^{2+}

ions are shown in red. **c**, The σ CD is shown together with the closely interacting regions of β and β' . **d**, Structural similarity between the *T. thermophilus* conserved σ CD, and the nonconserved regions of the *T. thermophilus* σ ND1 and *E. coli* σ 70 fragment.

helix may alternate between the flipped-out and normal α -helical conformations, depending on the position of the β' G loop (the 'trigger' loop).

Taken together, these structural data allow us to propose a possible mechanism for RNAP translocation during RNA synthesis. At the step of 'relaxation' (after translocation, before the next reaction), the template base at position $i+1$ is paired with the substrate nucleoside triphosphate (NTP), the bridge helix is in an all α -helical conformation, the Arg 1,096 bridges the i and $i+2$ DNA phosphates, and the flexible trigger loop is distal (rather than proximal) to the bridge α -helix. After phosphodiester bond formation, a signal induces the movement of the trigger loop towards the bridge α -helix, pushing out the 'switch' residues. In their flipped-out conformation, the switch residues may engage the DNA phosphate at position $i+1$ and bring the bridge α -helix under the DNA backbone towards the $i+2$ nucleotide. During this step, Arg 1,096 may also switch its interacting partner from DNA phosphates to the side chain of an acidic (polar) switch residue, thus simultaneously stabilizing the flipped-out conformation of the switch residues and facilitating the translocation of the enzyme (Fig. 5c, d).

Promoter recognition by the RNAP holoenzyme

A working model of bacterial open promoter complex has been constructed that is largely consistent with the genetic, biochemical and structural data²⁵. In this model, a straight dsDNA fragment is

placed to indicate the orientation of the promoter region upstream of its -10 element on the surface of the RNAP core enzyme. The presence of the σ subunit in our structure allows us to further improve on this model. After superposition of the core and holoenzyme structures, a simple translation of this dsDNA fragment is enough to bring it into plausible assembly with the holoenzyme (Fig. 6a). In this model, the -35 promoter hexamer is recognized by the recognition helix in Hth6 of the σ CD (region 4.2). The -10 promoter region apparently passes between the two central domains of σ , ND1 and ND2, which form a V-shaped nook at the entrance to the active site cleft. However, the 17-bp spacer between the -35 and -10 regions appears to be too long to be accommodated as a straight dsDNA fragment between the two promoter recognition modules of σ . Therefore, a local bend or a kink in the DNA seems necessary to bring the -10 promoter element and region 2.4 of σ into contact (Fig. 6a), making the model consistent with the biochemical and genetic data^{14,15,36,37}. In particular, in this model, the residues from regions 2.3–2.4 (σ 241–255) and 2.4–2.5 (σ 256–269) are inserted in the major groove of the DNA, in the vicinity of the -12 , -11 and -13 to -16 base pairs, respectively, which would allow base-specific recognition of the -10 and extended -10 regions of the promoter (Fig. 6a).

DNA melting by RNAP holoenzyme

After forming a closed complex at the promoter, the holoenzyme melts and unwinds the DNA near the -10 region (nt -12 to $+2$) to

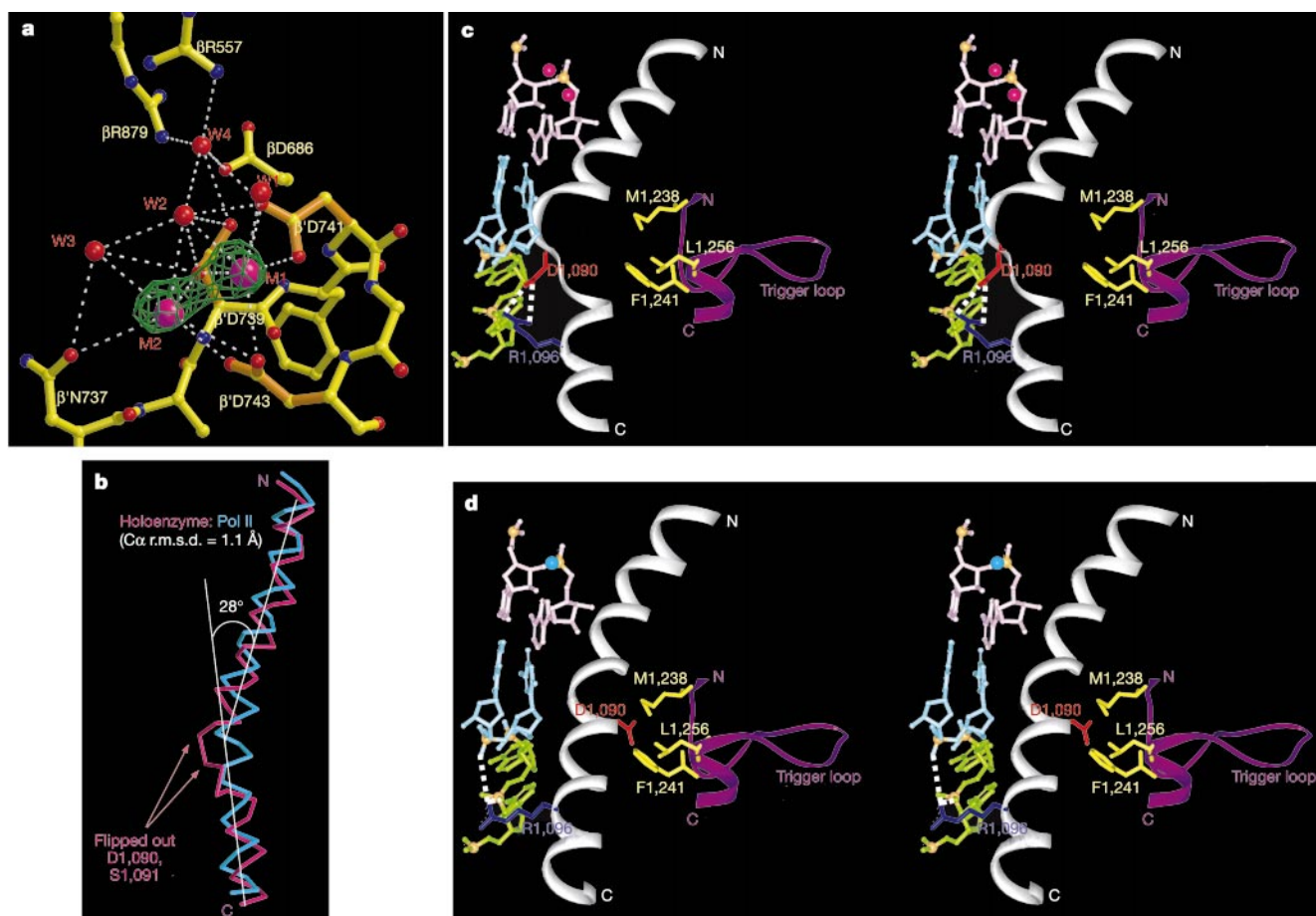


Figure 5 Catalytic centre. **a**, The Mg^{2+} -binding site. The slow annealing omit $|F_o - F_c|$ map contoured at 4.5σ (green) is shown for Mg^{2+} ions. Mg^{2+} ions (magenta) and water molecules (red) are shown as spheres. **b**, Superposition of the holoenzyme (magenta) and Pol II²³ (cyan) bridge helices. **c**, **d**, Two conformational states (stereo views) of the bridge

helix (white); in the holoenzyme (**c**) and in Pol II (**d**). The G-loop (purple), modelled for Pol II in **d**, and the DNA/RNA (cyan/pink) hybrid²³ (downstream DNA, green), modelled for the holoenzyme in **c**, are shown. The catalytic Mg^{2+} ions are shown as magenta (**c**) and cyan (**d**) spheres.

form an open promoter complex. The analysis of the holoenzyme structure and the model for the closed promoter complex discussed above allow us to hypothesize as to how the holoenzyme may catalyse promoter opening. Once the closed promoter complex is formed, the downstream DNA would eventually move towards its binding site on the RNAP, which was previously localized in a deep cavity^{23–25}, about 30 Å away from the location of the –10 recognition site (Fig. 6b, c). To achieve this, the dsDNA is likely to be kinked or bent near the –12 to –11 region of the promoter, in the immediate vicinity of the entrance to the active site cleft (Fig. 6b). Disruption of base-pairing at the –12 to –10 region and DNA strand separation (nucleation of promoter melting³⁸) should follow the specific recognition of this region by σ . We speculate that the melting of two or three DNA base pairs would sufficiently increase the DNA flexibility to direct the downstream dsDNA towards its binding site on the RNAP. However, the downstream dsDNA access to the binding site is still obstructed by the presence of the conserved σ α -helix (σ 74–92, region 1.2) and the β helix–loop–helix fragment (β 232–259) (Fig. 6b, c). The distance between these elements (~16 Å) is much shorter than the dsDNA diameter of about 22 Å (Fig. 6c). Interestingly, in our model, the β -subunit loop (β 242–253) protruding from the helix–loop–helix motif fits nicely in the dsDNA major groove, in close proximity to the +1 to +3 nt positions, which are known to be the end point of promoter melting (Fig. 6a, b). The sequence analysis shows a consensus motif of six residues—(L/M)RPGE—in bacterial RNAPs (β 242–247 in *T. thermophilus*), which are spatially located at the tip of the β loop. This β loop may serve as a gate that allows single-stranded, but not double-

stranded, DNA to go through. By interacting with DNA, this ‘gate’ loop may also stabilize the DNA orientation to favour melting, and/or to facilitate the downstream DNA rotation or oscillation around the helical axis to unwind the dsDNA. It should be noted that in the core enzyme, which is incapable of melting dsDNA, the absence of σ opens enough space at this region for the dsDNA to go through (Fig. 6d). After the DNA melting reaches the β loop (around nt +2 or +3), the unwound DNA and the downstream dsDNA may continue their movement towards the active site to complete the transcription bubble formation.

The mechanism of promoter DNA melting is largely unknown, and the ‘gate’ loop of the β subunit described here has never been the subject of genetic or biochemical studies. Future analyses of this region are required to shed light on its possible function.

Promoter escape

To better understand the mechanism of promoter escape by RNAP and σ release, we constructed a model of a DNA–RNA hybrid (9 bp) bound to the holoenzyme, based on a simple superposition of the conserved catalytic domains from the crystal structures of the eukaryotic elongation complex²³ and the *T. thermophilus* holoenzyme. According to our model, the DNA–RNA hybrid probably fits in a narrow passage (~11 Å wide) formed by the helix–loop–helix β fragment (β 375–408), and a hairpin loop from σ 70 LD (σ 318–329) protruding into the major groove of the DNA–RNA hybrid double helix (Fig. 6e). The putative interactions of the σ hairpin loop with the nucleotides in the hybrid may further deform the unusual ‘underwound’ conformation of the DNA–RNA

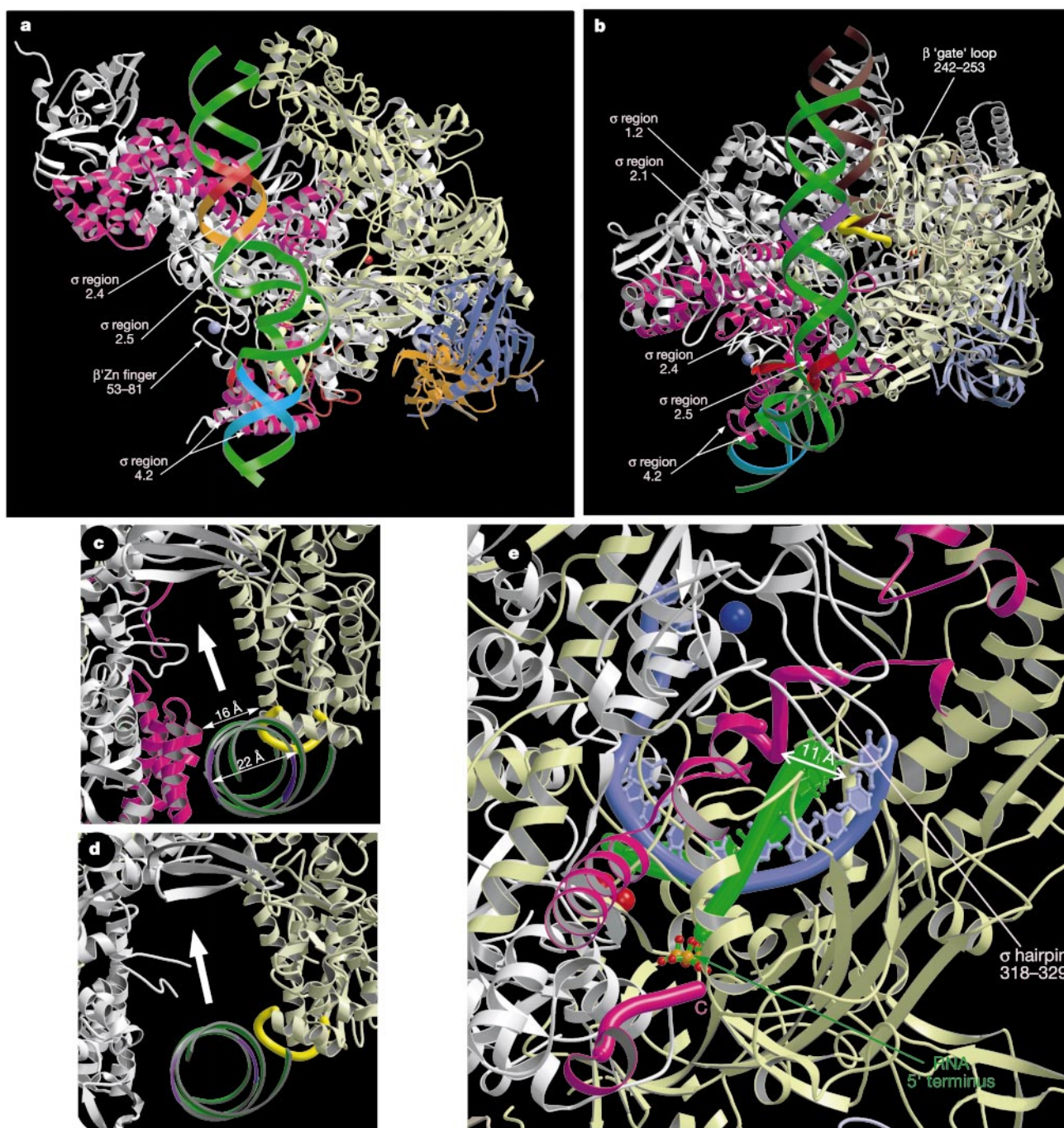


Figure 6 Models of the holoenzyme–nucleic acid complexes. The protein colour coding is the same as in Fig. 1c, except all of β' is shown in white. **a, b**, The closed (**a**) and partially melted (**b**) promoter complexes. **c, d**, The same DNA model as in **b**, but looking down the dsDNA axis for the holoenzyme (**c**) and for the *T. aquaticus* core (**d**). The dsDNA (green) contains the -35 (cyan), -10 (red), extended -10 (orange) and $+1$ to $+3$ (blue)

promoter regions. A brown dsDNA (**b**) indicates the downstream DNA-binding site in the open promoter complex. **e**, The model of the holoenzyme complex with the DNA/RNA (blue/green) hybrid and the ssRNA transcript. The Mg^{2+} (red) and Pb^{2+} (blue) ions are shown as spheres.

hybrid²³. Thus, the loop may act as a destabilizer to facilitate the disruption of the DNA–RNA hybrid base pairs. At the same time, the hybrid, competing with the σ destabilizer loop for limited space, may displace the loop from the active site cavity, thereby initiating the release of σ from the promoter and/or RNAP. This putative competition may also result in the release of abortive RNA products. After strand separation, the RNA chain is likely to enter the RNA-exit channel in an extended, rather than stacked, conformation. This is because the channel is substantially narrowed in the holoenzyme structure (10–12 Å in diameter, compared with 16–

19 Å in the *T. aquaticus* core) by the presence of the σ LD and the more closed conformation of the β flap domain. When RNAP synthesizes a 12-nt RNA, the 5'-terminal 3 or 4 nt in the extended conformation may easily reach the outlet of the RNA-exit channel, which in the holoenzyme structure is blocked by the σ C terminus (Fig. 6e). The potential steric collision between σ CD and nascent RNA may facilitate the displacement of the σ CD from the RNAP and from the -35 region of the promoter, to trigger the dissociation of $\sigma 70$ and promoter escape. This mechanism is consistent with recent biochemical data³⁹. □

Methods

Crystallization and data collection

The *T. thermophilus* holoenzyme was purified as described⁴⁰. The holoenzyme crystals were grown by the hanging-drop vapour-diffusion technique. For crystallization, 2 µl of well solution were mixed with 2 µl of protein solution (protein concentration 10 mg ml⁻¹). The final conditions, 13 mM magnesium formate, 5% polyethylene glycol 400, 2 mM spermine, 2 mM dithiothreitol and 20 mM MES buffer (pH 5.8), yielded high-quality hexagonal crystals⁴⁰. The crystals of the native enzyme and several heavy-atom derivatives were subjected to data collection at SPring-8 synchrotron beam lines. Among these, the crystal of the lead heavy-atom derivative—(CH₃)₃Pb(CH₃COO)₂—showed the best diffraction pattern, and the data from this crystal were used for refinement (see Supplementary Information). To increase the data completeness, data from the same crystal were collected at the BL44XU and BL45XU beam lines. The data were processed by the HKL2000 program package⁴¹.

Structure determination and refinement

Careful analysis of the diffraction data, as described previously⁴⁰, revealed that the holoenzyme crystals belong to the *P*₃2 space group, have perfect merohedral twinning along a three-fold axis⁴², and contain two molecules in the asymmetric unit, the orientation of which closely resembles crystallographic *P*₆₅ symmetry. The molecular replacement solution, which was obtained previously for the 3 Å native data using the *T. aquaticus* RNAP core enzyme structure (PDB number 1I6V) as a search model⁴⁰, was used as a starting point for the holoenzyme structure determination. The rigid body refinement using the CNS program⁴³ converged to an *R*_{free} of 39.2% at 2.6 Å resolution. At this stage, the solvent flipping procedure⁴⁴, as implemented in the CNS program⁴³, greatly improved the electron density map of the σ subunit and the regions of the core enzyme that were absent from the *T. aquaticus* protein structure. The refinement and model building revealed several 'local' mistakes in the sequence assignment of the *T. aquaticus* core, with the largest misassigned region of 70 amino acids (β460–530) in the β subunit. The model was gradually improved by alternate cycles of CNS refinement and model building using the O program⁴⁵. The two lead ions in each holoenzyme molecule substitute for zinc in the two zinc-binding sites. This is revealed by the more than 20σ peaks in the difference electron density map at 3.5 Å resolution produced between the lead derivative and native data (data not shown). At the final stage, water molecules were added to the model, using alternate steps of the 'water pick' and 'water delete' procedures as implemented in the CNS program⁴³. The refinement converged to a final *R*-factor value of 22.8% (*R*_{free} = 27.4%) for all data at 2.6 Å resolution (see Supplementary Information). The asymmetric unit contains two holoenzyme molecules, with r.m.s.d. = 1.05 Å over all main-chain atoms. Figures 1, 2, 5a and 6 were prepared using the programs O⁴⁵, Molscript⁴⁶, Bobscript⁴⁷ and Raster3D⁴⁸. Other figures were prepared by the WebLab ViewerPro program (MSI).

Received 26 February; accepted 25 April 2002; doi:10.1038/nature752.

Published online 8 May 2002.

- Sweetser, D., Nonet, M. & Young, R. A. Prokaryotic and eukaryotic RNA polymerase have homologous core subunits. *Proc. Natl Acad. Sci. USA* **84**, 1192–1196 (1987).
- Ebright, R. H. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* **304**, 687–698 (2000).
- Burgess, R. R., Travers, A. A., Dunn, J. J. & Bautz, E. K. Factors stimulating transcription by RNA polymerase. *Nature* **221**, 43–44 (1969).
- Gross, C., Lonetto, M. & Losick, R. in *Transcriptional Regulation* (eds McKnight, S. R. & Yamamoto, K. R.) 129–176 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1992).
- Record, M. T. J., Reznikoff, W., Craig, M., McQuade, K. & Schlax, P. in *Escherichia coli and Salmonella* (ed. Neidhart, F. C.) 792–820 (ASM, Washington DC, 1996).
- von Hippel, P. An integrated model of the transcription complex in elongation, termination, and editing. *Science* **281**, 660–665 (1998).
- Bar-Nahum, G. & Nudler, E. Isolation and characterization of σ^{70} -retaining transcription elongation complexes from *Escherichia coli*. *Cell* **106**, 443–451 (2001).
- Mukhopadhyay, K. *et al.* Translocation of σ^{70} with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* **106**, 453–463 (2001).
- Craig, M. L. *et al.* DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase–promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. *J. Mol. Biol.* **283**, 741–756 (1998).
- Gross, C. *et al.* The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harbor Symp. Quant. Biol.* **63**, 141–155 (1998).
- Ishihama, A. Functional modulation of *Escherichia coli* RNA polymerase. *Annu. Rev. Microbiol.* **54**, 499–518 (2000).
- Lonetto, M., Gribskov, M. & Gross, C. The σ^{70} family: sequence conservation and evolutionary relationships. *J. Bacteriol.* **174**, 3843–3849 (1992).
- Dombroski, A. J., Walter, W. A., Record, M. T., Siegel, D. A. & Gross, C. A. Polypeptides containing highly conserved region of transcription initiation factor Σ 70 exhibit specificity of binding to promoter DNA. *Cell* **70**, 501–512 (1992).
- Barne, K. A., Bown, J. A., Busby, S. J. W. & Minchin, S. D. Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended -10' motif at promoters. *EMBO J.* **16**, 4034–4040 (1997).
- Fenton, M. S., Lee, S. J. & Gralla, J. D. *Escherichia coli* promoter opening and -10 recognition: mutational analysis of σ^{70} . *EMBO J.* **19**, 1130–1137 (2000).
- Marr, M. T. & Roberts, J. W. Promoter recognition as measured by binding of polymerase to non-template strand oligonucleotide. *Science* **276**, 1258–1260 (1997).
- Huang, X., Lopez de Saro, F. J. & Helmman, J. D. Sigma factor mutations affecting the sequence-selective interaction of RNA polymerase with -10 region single-stranded DNA. *Nucleic Acids Res.* **25**, 2603–2609 (1997).
- Sharp, M. *et al.* The interface of σ with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes Dev.* **13**, 3015–3026 (1999).
- Lesley, S. A. & Burgess, R. R. Characterization of the *Escherichia coli* transcription factor sigma 70: localization of a region involved in the interaction with core RNA polymerase. *Biochemistry* **28**, 7728–7734 (1989).
- Owens, J. T. *et al.* Mapping the σ^{70} subunit contact sites on *Escherichia coli* RNA polymerase with a σ^{70} -conjugated chemical protease. *Proc. Natl Acad. Sci. USA* **95**, 6021–6026 (1998).
- Zhang, G. *et al.* A crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
- Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292**, 1863–1876 (2001).
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
- Korzheva, N. *et al.* A structural model of transcription elongation. *Science* **289**, 619–625 (2000).
- Naryshkin, N., Revyakina, A., Kim, Y., Mekler, V. & Ebright, R. H. Structural organization of the RNA polymerase–promoter open complex. *Cell* **101**, 601–611 (2000).
- Coulombe, B. & Burton, Z. F. DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. *Microbiol. Mol. Biol. Rev.* **63**, 457–478 (1999).
- Dombroski, A. J., Walter, W. A. & Gross, C. A. Amino-terminal amino acids modulate σ -factor DNA-binding activity. *Genes Dev.* **7**, 2446–2455 (1993).
- Siegel, D. A., Hu, J. C., Walter, W. A. & Gross, C. A. Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **206**, 591–603 (1989).
- Gardella, T., Moyle, H. & Susskind, M. M. A mutant *Escherichia coli* sigma 70 subunit of RNA polymerase with altered promoter specificity. *J. Mol. Biol.* **206**, 579–590 (1989).
- Malhotra, A., Severinova, E. & Darst, S. A. Crystal structure of a σ 70 subunit fragment from *E. coli* RNA polymerase. *Cell* **87**, 127–136 (1996).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- Arthur, T. M., Anthony, L. C. & Burgess, R. R. Mutational analysis of $\beta'_{260-309}$, a σ^{70} binding site located on *Escherichia coli* core RNA polymerase. *J. Biol. Chem.* **275**, 23113–23119 (2000).
- Young, B. A. *et al.* A coiled-coil from the RNA polymerase β' subunit allosterically induces selective non-template strand binding by σ^{70} . *Cell* **105**, 935–944 (2001).
- Dieci, G. *et al.* A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *EMBO J.* **14**, 3766–3776 (1995).
- Brautigam, C. A. & Steitz, T. A. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **1**, 54–63 (1998).
- Bown, J. A. *et al.* Organization of open complexes at *Escherichia coli* promoters. *J. Biol. Chem.* **274**, 2263–2270 (1999).
- Owens, J. T. *et al.* Mapping the promoter DNA sites proximal to conserved regions of σ^{70} in an *Escherichia coli* RNA polymerase–*lacUV5* open promoter complex. *Biochemistry* **37**, 7670–7675 (1998).
- Lim, H. M., Lee, H. J., Roy, S. & Adhya, S. A "master" in base unpairing during isomerization of a promoter upon RNA polymerase binding. *Proc. Natl Acad. Sci. USA* **98**, 14849–14852 (2001).
- Daube, S. S. & von Hippel, P. H. Interactions of *Escherichia coli* σ^{70} within the transcription elongation complex. *Proc. Natl Acad. Sci. USA* **96**, 8390–8395 (1999).
- Vassilyeva, M. N. *et al.* Purification, crystallization and initial crystallographic analysis of RNA polymerase holoenzyme from *Thermus thermophilus*. *Acta Crystallogr.* (submitted).
- Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Yeates, T. D. Detecting and overcoming crystal twinning. *Methods Enzymol.* **276**, 344–358 (1997).
- Brünger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**, 621–628 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
- Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank M. Yamamoto for assistance during the data collection at the SPring-8 synchrotron beam line, BL45. We are grateful to T. Yeates for discussions and advice concerning the merohedral twinning problem. This work was supported in part by a grant from the National Institutes of Health to S.B. and by a grant from the Organized Research Combination System of Science and Technology Agency (Japan) to S.Y.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.G.V. (e-mail: dmitry@yumiyoishi.harima.riken.go.jp), S.Y. (e-mail: yokoyama@biochem.s.u-tokyo.ac.jp) or S.B. (e-mail: serbor@asan.com). The atomic coordinates have been deposited in the Protein Data Bank under accession number 1IW7.

The recent breakup of an asteroid in the main-belt region

David Nesvorný, William F. Bottke Jr, Luke Dones & Harold F. Levison

Southwest Research Institute, 1050 Walnut St, Suite 426, Boulder, Colorado 80302, USA

The present population of asteroids in the main belt is largely the result of many past collisions^{1,2}. Ideally, the asteroid fragments resulting from each impact event could help us understand the large-scale collisions that shaped the planets during early epochs^{3–5}. Most known asteroid fragment families, however, are very old and have therefore undergone significant collisional and dynamical evolution since their formation⁶. This evolution has masked the properties of the original collisions. Here we report the discovery of a family of asteroids that formed in a disruption event only 5.8 ± 0.2 million years ago, and which has subsequently undergone little dynamical and collisional evolution^{6,7}. We identified 39 fragments, two of which are large and comparable in size (diameters of ~ 19 and ~ 14 km), with the remainder exhibiting a continuum of sizes in the range 2–7 km. The low measured ejection velocities suggest that gravitational re-accumulation after a collision may be a common feature of asteroid evolution. Moreover, these data can be used to check numerical models of larger-scale collisions⁸.

Up to now, ejecta from a few tens of major collisions (that is, asteroid families) have been observed in the main belt^{9,10}. To identify an asteroid family, researchers look for clusters of asteroid positions in the space of so-called proper orbital elements: the proper semimajor axis (a_p), proper eccentricity (e_p) and proper inclination (i_p). The orbital elements describe the size, shape and tilt of orbits. Proper orbital elements, being more constant over time than instantaneous orbital elements¹¹, provide a dynamical criterion of whether or not a group of bodies has a common ancestor. Unfortunately, the observed asteroid families are old enough (that is, hundreds of millions to billions of years old^{6,12}) to have been substantially eroded and dispersed by: (1) secondary collisions⁵, (2) chaotic orbital evolution^{12,13}, and (3) semimajor axis mobility due to radiation effects⁷. These effects make it difficult to determine the conditions that existed immediately after the family breakup event and the age of the family itself.

Fragments produced by recent collisions, on the other hand, would suffer little erosion in the interim following a breakup event, and thus would provide better examples for understanding impact physics than older families. To search for recently formed asteroid families, we applied a cluster-detection algorithm called the hierarchical clustering method⁹ (HCM) to a proper-element database^{11,14}. The output of the HCM algorithm is a number of clusters in the (a_p , e_p , i_p) space. Each cluster consists of member bodies linked by a 'chain' in the (a_p , e_p , i_p) space, with the length of each link, d (see Methods section), corresponding to a proper-element difference not exceeding d_{cutoff} . In order to detect signatures of compact, and presumably young, collisional clusters, we used $d_{\text{cutoff}} = 10 \text{ m s}^{-1}$. This value is an order of magnitude less than what is usually used for asteroid family searches.

The HCM algorithm found nine clusters with five or more member asteroids. Among them, four clusters form a prominent compact feature of 39 bodies within the region known as the Koronis family (Fig. 1). At $d_{\text{cutoff}} = 15 \text{ m s}^{-1}$, the four clusters join together, but also agglomerate a dozen additional asteroids, forming a similar but more loosely connected structure. In this work, we analyse the orbital and size-frequency distribution of the 39 asteroids detected at $d_{\text{cutoff}} = 10 \text{ m s}^{-1}$. This 39-body cluster is statistically significant at greater than the 99% level

(see Methods section).

The orbital distribution of the 39-body cluster is diagonally shaped in (a_p , e_p), and appears to fit inside one of the similarly shaped 'equivelocity' ellipses shown in Fig. 1. To create these ellipses, we launched test bodies isotropically from the centre of the cluster (presumably the impact site) at a selected velocity impulse δV . Using Gauss's equations (see, for example, ref. 14), we then computed the change in their proper elements (δa_p , δe_p , δi_p). To determine the size, shape and orientation of the ellipses, we experimented with various values of V_{max} , the maximum velocity among our test bodies, the true anomaly f of the parent body (that is, the angle between the parent body's location and the perihelion of its orbit) and the parent body's perihelion argument ω at the instant of the impact (that is, the angle between the perihelion and the ascending node).

We found that ellipsoids having $V_{\text{max}} \approx 15 \text{ m s}^{-1}$, $-30^\circ \leq f \leq 30^\circ$ and $-45^\circ \leq \omega + f \leq 45^\circ$ provide good fits to our 39-body cluster (Fig. 1). These results indicate that the disruption event occurred while the parent body was near perihelion, and that the fragments were ejected in nearly isotropic directions.

No other known asteroid family shows a comparable fit to that shown in Fig. 1a. Instead, most observed families have box-like shapes in (a_p , e_p , i_p) space. We believe these observed family configurations are by-products of long-term dynamical evolution^{7,13} rather than the result of an unusual δV distribution. Similarly, our numerical experiments suggest that the structure of the 39-body cluster in (a_p , e_p , i_p) space cannot be older than a few times 10^7 years, or it would have been erased by asteroid mobility

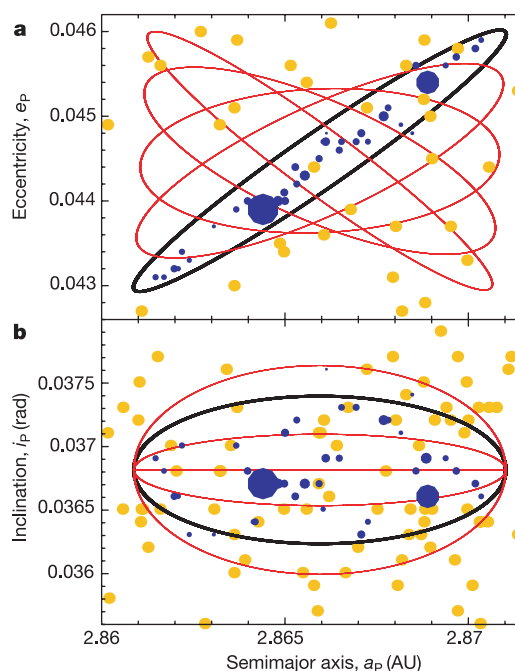


Figure 1 The asteroid family in orbital phase space. The orbital structure of the identified cluster is compatible with a recent asteroid breakup at its location: **a**, a_p and e_p ; **b**, a_p and i_p . The size of each blue symbol is proportional to the diameter of a body. Gold dots indicate the background bodies in the Koronis family. The equivelocity ellipses were computed from Gauss's equations. The red ellipses correspond to isotropic ejection fields with $\delta V = 15 \text{ m s}^{-1}$ and different values of f and ω (see text). In **a**, $f = 30^\circ$ (black), 60° , 90° , 120° and 150° are shown. In **b**, the equivelocity ellipses were computed for $\omega + f = 0^\circ$, 45° (black), 70° and 90° . The diagonal shape of the 39-body cluster in the (a_p , e_p) plane suggests that the collision occurred near the perihelion of the parent's body orbit. The overall structure in (a_p , e_p , i_p) provides a reasonable match to the ellipses with parameters $\delta V \leq 15 \text{ m s}^{-1}$, $-30^\circ \leq f \leq 30^\circ$ and $-45^\circ \leq \omega + f \leq 45^\circ$.

mechanisms^{7,13}. Thus, we conclude that the 39-body cluster must be young—so young, in fact, that it may be the first essentially unaltered disruption structure ever identified in the main belt.

To determine the age of the cluster, we numerically integrated the orbits of our 39 bodies into the past. Our goal was to show that in some previous epoch, the heliocentric orbits of all cluster members were nearly the same. As $\delta V \ll V_{\text{orb}}$, where V_{orb} is the orbital velocity of cluster members, we expect that this conjunction of orbital elements should have occurred shortly after the parent body disrupted.

There are two angles that determine the orientation of an orbit in space: the longitude of the ascending node (Ω), and the argument of perihelion (ω). Owing to planetary perturbations, these angles evolve with different but nearly constant speeds for individual asteroid orbits (for example, precession periods of $\sim 19,000$ yr and $9,500$ yr, respectively, for cluster members). Today, the orbits of the cluster members are oriented differently in space because their slightly dissimilar periods of Ω and ω produce slow differential rotation of their orbits with respect to each other. Eventually, this effect allows Ω and ω to obtain nearly uniform distributions in $[0^\circ, 360^\circ]$. For a short time after the parent body breakup, however, the orientations of the fragments' orbits must have been nearly the same.

To deal solely with well-determined orbits, we used the orbital elements of the 13 cluster members that have a number designation¹⁶. The orbits were propagated backward in time using a symplectic integration scheme (the Wisdom–Holman map in the software package Swift^{17,18}) with a 0.01-year step size. The inte-

gration accounted for the gravitational effect of the four giant planets, but neglected the effect of the inner planets, radiation forces and relativistic corrections. Given the small eccentricities and inclinations of the cluster, its location in a regular part of the main belt and the short integration interval, the latter effects can be ignored. We nevertheless verified that the inclusion of the inner planets in the integration did not change our results.

We find that about 5.8 Myr ago, all integrated orbits had nearly the same orientation in space (Fig. 2). This convergence did not happen by chance; we estimate that the probability that 13 unrelated orbits would randomly meet at one time over the age of the Solar System is $< 10^{-6}$. We also tested this idea by integrating their orbits forwards and backwards by 50 Myr. No other alignment was found. Thus, we conclude that the cluster members had a common collisional origin about 5.8 Myr ago.

To roughly estimate the error on this age (Δt), we did the following. First, we computed the angular distance between the cluster members' ascending nodes at various times ($t = -5.8 \pm \Delta t$). Next, we averaged these values for each t . The minimum average value, found at $t = -5.8$ Myr, was about 10° . For $\Delta t = 0.2$ Myr, the average angular distance had doubled to 20° . Thus, we estimate that the breakup event occurred 5.8 ± 0.2 Myr ago.

The size–frequency distribution of the 39-body cluster shows two large bodies, (832) Karin and (4507) 1990 FV, with absolute magnitudes $H = 11.2$ (diameter $D \approx 19$ km for a 16% albedo) and 11.8 ($D \approx 14.4$ km), respectively. As (832) Karin is the largest body in the observed population, we adopt the name 'Karin cluster'. The Karin cluster's size–frequency distribution shows 6, 15, 11, 3 and 2 bodies in 0.5-mag bins between $H = 13.5$ mag and 16.0 mag ($D \approx 6.6$ – 2.1 km). The size–frequency distribution of this continuum is strongly affected by observational incompleteness.

To compute the size of the parent body of the Karin cluster, we need to account for its undetected members. Using a reasonable estimate of the observational incompleteness of the main belt for asteroids with $H < 16.0$ mag (ref. 19), we infer that the diameter of the parent body was $D_{\text{PB}} \approx 24.5 \pm 1$ km. D_{PB} corresponds to a spherical body with volume equal to the estimated total volume of the bodies with $H < 16.0$ mag. It is likely that the actual size of the parent body was larger than D_{PB} because the contribution of the population at $H > 16.0$ mag was not used. According to the above estimate, (832) Karin contains $47 \pm 6\%$ of the mass of the parent body. The estimated disruption rate of asteroids in the main belt with $D \approx 25$ km (1 per ~ 5 Myr) is consistent with our derived age for the Karin cluster (D. Durda, personal communication).

According to numerical models⁸, an impact at 5 km s^{-1} of a ≈ 3 -km-diameter projectile on the 24.5-km-diameter parent body of the Karin cluster produces a fragment comparable to the size of (832) Karin. The next largest Karin cluster member, however, is only $\sim 25\%$ smaller than (832) Karin itself. Because impacts onto solid homogeneous parent bodies are thought to yield large fragments with significant diameter differences², these measurements suggest that either the parent body was fractured before the breakup event or the parent body had a heterogeneous interior structure.

Another possibility is that low ejection velocities allowed gravitational reaccumulation to modify the sizes of the largest fragments. Assuming $f = \pm 30^\circ$ and $\omega + f = \pm 45^\circ$, we invert Gauss's equations and find that typical δV values are $\lesssim 15 \text{ m s}^{-1}$. We also find a correlation between the calculated δV values and the size of a body, with $\delta V \propto D^{-1}$. These ejection velocities are comparable to the escape velocity of a spherical, 25-km-diameter body with a bulk density of 2.5 g cm^{-3} (that is, $V_{\text{esc}} \approx 15 \text{ m s}^{-1}$). We expect that the largest cluster members will agglomerate fragments having ejection velocities $V_{\text{ejc}} < V_{\text{esc}}$, assuming reasonable distributions of V_{ejc} (ref. 20).

The relative youth and known age of the members of the Karin cluster may help us answer several important questions in asteroid

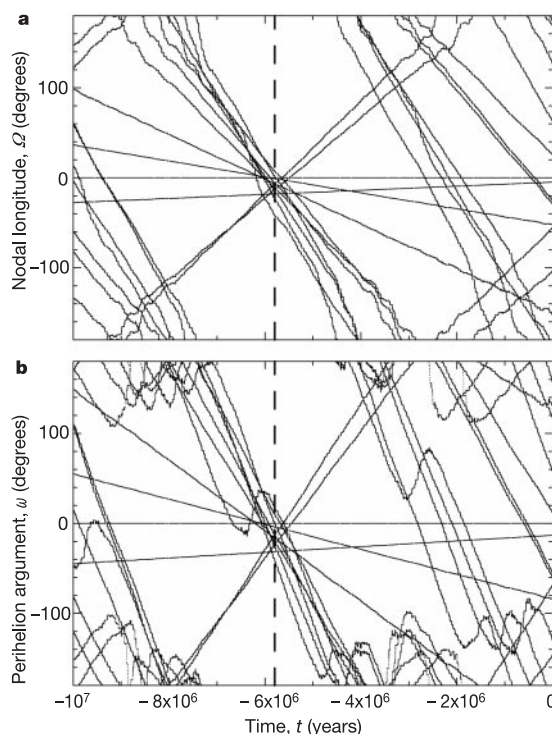


Figure 2 The convergence of angles at about 5.8 Myr ago strongly suggests that the identified cluster was created by a recent catastrophic collision. The plot shows past orbital histories of the 13 numbered members of the 39-body cluster: **a**, nodal longitude Ω , and **b**, perihelion argument ω . Values relative to (832) Karin are shown. For clarity, the evolutions were smoothed by a 0.5-Myr running time window. Ω and ω have small oscillations about the average values. At $t = -5.8$ Myr (broken vertical line), the nodal longitudes and perihelion arguments of all 13 asteroids were nearly the same. Thus, at $t = -5.8$ Myr, all orbits were nearly identical.

geology and impact physics. Recent work has suggested that S-type asteroids with 'fresh' surfaces may have spectra that resemble (or trend towards) ordinary chondrite spectra (see, for example, ref. 21). Because Karin cluster asteroids are S-type²² and have probably experienced minimal space weathering²³, we believe they could be used to examine this possibility. It would also be useful to compare the spectra of various Karin cluster asteroids, all of which should share the same age (that is, 5.8 Myr). Similarities and differences might help us to determine the rate of space weathering among S-type main-belt asteroids, and whether spectral alterations by space weathering are a function of time alone or are also dependent on asteroid size. Moreover, if the surfaces of the Karin-cluster members were given fresh surfaces by impact 5.8 Myr ago, the craters formed since that time by hypervelocity impacts may be used to infer the current crater production rate in the main belt, and the unknown shape of the main-belt's size distribution at small asteroid sizes.

It is possible that the Karin cluster may be the source of some asteroidal material evolving towards Earth. For example, we believe that impacts on Karin cluster asteroids could be the source of the zodiacal dust β -band discovered by the IRAS satellite²⁴. We base this idea on two pieces of evidence: (1) the Karin cluster and the model-derived source of the β -band both have mean inclinations near 2.1° , and (2) the narrow inclination span of the β -band source ($\Delta i \approx 0.09^\circ$) is analogous to that of our compact 39-body cluster ($\Delta i \approx 0.08^\circ$)²⁵. (The inclination span of the Koronis family, which has generally been taken as the source of the β -band, is $\Delta i \approx 0.45^\circ$). As a second example, we believe the Karin breakup event could have produced some meteorites. These putative objects would need to have compositions consistent with S-type asteroids and cosmic-ray exposure²⁶ ages $\lesssim 5.8$ Myr. $\square$

Methods

Hierarchical clustering method

The hierarchical clustering method (HCM) starts with an individual asteroid position in the proper elements space, and identifies bodies in its neighbourhood with mutual distances less than a threshold limit (d_{cutoff}). We define the distance in the (a_p, e_p, i_p) space by:

$$d = na_p \sqrt{C_a(\delta a_p/a_p)^2 + C_e(\delta e_p)^2 + C_i(\delta \sin i_p)^2} \quad (1)$$

where na_p is the heliocentric velocity of an asteroid on a circular orbit having the semimajor axis a_p . $\delta a_p = |a_p^{(1)} - a_p^{(2)}|$, $\delta e_p = |e_p^{(1)} - e_p^{(2)}|$, and $\delta \sin i_p = |\sin i_p^{(1)} - \sin i_p^{(2)}|$. The indices (1) and (2) denote the two bodies under consideration. C_a , C_e and C_i are constants: we use $C_a = 5/4$, $C_e = 2$ and $C_i = 2$ (ref. 10). Other choices of these constants found in the literature yield similar results.

Statistical significance of the 39-body cluster

To demonstrate the >99% statistical significance of the 39-body cluster, we generated 100 synthetic orbital distributions corresponding to the Koronis family determined at $d_{\text{cutoff}} = 60 \text{ m s}^{-1}$ (that is, 1,500 asteroid positions at $2.83 < a_p < 2.95 \text{ AU}$, $0.04 < e_p < 0.06$ and $0.033 < i_p < 0.04$), and applied our HCM algorithm to these data. Applying $d_{\text{cutoff}} = 10 \text{ m s}^{-1}$, we were unable to find a cluster containing more than five members. We also used the HCM algorithm on 100 computer-generated asteroid belts (that is, 66,000 random orbital positions at $2.1 < a_p < 3.25 \text{ AU}$, $e_p < 0.3$ and $i_p < 0.3$). Once again, $d_{\text{cutoff}} = 10 \text{ m s}^{-1}$ yielded no meaningful structures.

Received 27 February; accepted 11 April 2002; doi:10.1038/nature00789.

1. Durda, D. D., Greenberg, R. & Jedicke, R. Collisional models and scaling laws: a new interpretation of the shape of the main-belt asteroid size distribution. *Icarus* **135**, 431–440 (1998).
2. Michel, P., Benz, W., Tanga, P. & Richardson, D. C. Collisions and gravitational reaccumulation: Forming asteroid families and satellites. *Science* **294**, 1696–1700 (2001).
3. Zappalà, V., Cellino, A., Dell'Oro, A. & Paolucci, P. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolucci, P. & Binzel, R.) (Univ. Arizona Press, in the press).
4. Chambers, J. E. & Wetherill, G. W. Making the terrestrial planets: N-body integrations of planetary embryos in three dimensions. *Icarus* **136**, 304–327 (1998).
5. Canup, R. M. & Asphaug, E. Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* **412**, 708–712 (2001).
6. Marzari, F., Davis, D. & Vanzani, V. Collisional evolution of asteroid families. *Icarus* **113**, 168–187 (1995).
7. Bottke, W. F., Vokrouhlický, D., Brož, M., Nesvorný, D. & Morbidelli, A. Dynamical spreading of asteroid families by the Yarkovsky effect. *Science* **294**, 1693–1696 (2001).
8. Benz, W. & Asphaug, E. Catastrophic disruptions revisited. *Icarus* **142**, 5–20 (1999).
9. Hirayama, K. Groups of asteroids probably by common origin. *Astron. J.* **31**, 185–188 (1918).
10. Zappalà, V., Cellino, A., Farinella, P. & Milani, A. Asteroid families. II. Extension to unnumbered multiopposition asteroids. *Astron. J.* **107**, 772–801 (1994).
11. Milani, A. & Knežević, Z. Asteroid proper elements and the dynamical structure of the asteroid main

belt. *Icarus* **107**, 219–254 (1994).

12. Milani, A. & Farinella, P. The age of the Veritas asteroid family deduced by chaotic chronology. *Nature* **370**, 40–41 (1994).
13. Nesvorný, D., Morbidelli, A., Vokrouhlický, D., Bottke, W. F. & Brož, M. The Flora family: a case of the dynamically dispersed collisional swarm? *Icarus* **157**, 155–172 (2002).
14. *Asteroids Dynamic Site* (<http://hamilton.dm.unipi.it/cgi-bin/astdys/astibo>) (7 May 2002).
15. Morbidelli, A., Zappalà, A., Moons, M., Cellino, A. & Gonczi, R. Asteroid families close to mean motion resonances: Dynamical effect and physical implications. *Icarus* **118**, 132–154 (1995).
16. IAU: Minor Planet Centre (<http://cfa-www.harvard.edu/cfa/ps/mpc.html>) (7 May 2002).
17. Levison, H. F. & Duncan, M. The long term dynamical behaviour of short-period comets. *Icarus* **108**, 18–36 (1994).
18. Wisdom, J. & Holman, M. Symplectic maps for the n-body problem. *Astron. J.* **102**, 1528–1538 (1991).
19. Jedicke, R. & Metcalfe, T. S. The orbital and absolute magnitude distributions of main belt asteroids. *Icarus* **131**, 245–260 (1998).
20. Fujiwara, A. et al. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 240–265 (Univ. Arizona Press, Tucson, 1989).
21. Binzel, R. P., Bus, S. J., Burbine, T. H. & Sunshine, J. M. Spectral properties of near-earth asteroids: Evidence for sources of ordinary chondrite meteorites. *Science* **273**, 946–948 (1996).
22. (<http://pdssbn.astro.umd.edu/SBNast/holdings/EAR-A-5-DDR-UBV-MEAN-VALUES-V1.0.html>) (6 Feb. 2001).
23. Clark, B. E., Hapke, B., Pieters, C. & Britt, D. in *Asteroids III* (eds Bottke, W., Cellino, A., Paolucci, P. & Binzel, R. P.) (Univ. Arizona Press, in the press).
24. Dermott, S. F., Nicholson, P. D., Burns, J. A. & Houck, J. R. Origin of the solar system dust bands discovered by IRAS. *Nature* **312**, 505–509 (1984).
25. Grogan, K., Dermott, S. F. & Durda, D. D. The size-frequency distribution of the zodiacal cloud: Evidence from the solar system dust bands. *Icarus* **152**, 251–267 (2001).
26. Marti, K. & Graf, T. Cosmic-ray exposure history of ordinary chondrites. *Annu. Rev. Earth Planet. Sci.* **20**, 221–243 (1992).

Acknowledgements

We thank R. Binzel, C. Chapman, D. Durda, O. Eugster, B. Gladman, D. Hamilton, R. Jedicke, A. Morbidelli, F. Namouni and M. Sykes for their suggestions.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.N. (e-mail: davidn@boulder.swri.edu).

Coulomb blockade and the Kondo effect in single-atom transistors

Jiwoong Park^{*†‡}, Abhay N. Pasupathy^{*‡}, Jonas I. Goldsmith[§], Connie Chang^{*}, Yuval Yaish^{*}, Jason R. Petta^{*}, Marie Rinkoski^{*}, James P. Sethna^{*}, Héctor D. Abruña[§], Paul L. McEuen^{*} & Daniel C. Ralph^{*}

^{*} Laboratory of Atomic and Solid State Physics; and [§] Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA

[†] Department of Physics, University of California, Berkeley, California 94720, USA

[‡] These authors contributed equally to this work

Using molecules as electronic components is a powerful new direction in the science and technology of nanometre-scale systems¹. Experiments to date have examined a multitude of molecules conducting in parallel^{2,3}, or, in some cases, transport through single molecules. The latter includes molecules probed in a two-terminal geometry using mechanically controlled break junctions^{4,5} or scanning probes^{6,7} as well as three-terminal single-molecule transistors made from carbon nanotubes⁸, C₆₀ molecules⁹, and conjugated molecules diluted in a less-conducting molecular layer¹⁰. The ultimate limit would be a device where electrons hop on to, and off from, a single atom between two contacts. Here we describe transistors incorporating a transition-metal complex designed so that electron transport occurs through well-defined charge states of a single atom. We examine two related molecules containing a Co ion bonded to polypyridyl ligands, attached to insulating tethers of different lengths. Chan-

ging the length of the insulating tether alters the coupling of the ion to the electrodes, enabling the fabrication of devices that exhibit either single-electron phenomena, such as Coulomb blockade, or the Kondo effect.

The molecules that we have investigated are depicted in Fig. 1a. They are coordination complexes in which one Co ion is bonded within an approximately octahedral environment to two terpyridinyl linker molecules with thiol end groups, which confer high adsorbability onto gold surfaces. The two molecules ($[\text{Co}(\text{tpy}-(\text{CH}_2)_5\text{-SH})_2]^{2+}$ and $[\text{Co}(\text{tpy-SH})_2]^{2+}$) differ by a five-carbon alkyl chain within the linker molecules (see Methods for details). These molecules were selected because it is known from electrochemical studies that the charge state of the Co ion can be changed from 2+ to 3+ at low energy. A cyclic voltammogram¹¹ for $[\text{Co}(\text{tpy-SH})_2]^{2+}$ adsorbed on a gold electrode in an acetonitrile/supporting electrolyte solution is shown in Fig. 1b, indicating that a positive voltage $V_s \approx +0.25$ V (measured against an Ag/AgCl reference) applied to the solution removes one electron from the ion. Similar results were obtained for $[\text{Co}(\text{tpy}-(\text{CH}_2)_5\text{-SH})_2]^{2+}$ (ref. 12).

Preparation of the transistors (schematically shown in Fig. 1c) begins with the thermal growth of a 30-nm SiO_2 insulating layer on top of a degenerately doped Si substrate used as a back gate. Continuous gold wires with widths of less than 200 nm, lengths of 200–400 nm and thicknesses of 10–15 nm are fabricated on the SiO_2 layer by electron beam lithography. The wires are cleaned with acetone, methylene chloride and oxygen plasma, and placed in a dilute solution of the molecules in acetonitrile for a day or more in order to form a self-assembled monolayer on the Au electrodes. The

wires coated with molecules are then broken by electromigration, by ramping to large voltages (typically over 0.5 V) at cryogenic temperatures while monitoring the current until only a tunnelling signal is present¹³. This produces a gap about 1–2-nm-wide, across which a molecule is often found. Electrical characteristics of the molecule are determined by acquiring current versus bias voltage (I - V) curves while changing the gate voltage (V_g).

First we discuss the results obtained for the longer molecule, $[\text{Co}(\text{tpy}-(\text{CH}_2)_5\text{-SH})_2]$. The measurements were performed in a dilution refrigerator with an electron temperature of less than 100 mK. In about 10% of 400 broken wires we see I - V curves as shown in Fig. 1c. The current is strongly suppressed up to some threshold voltage that depends on V_g , and then it increases in steps. In Fig. 2 we show higher-resolution colour-scale plots of the differential conductance $\partial I/\partial V$ at low bias, as a function of V and V_g for three different devices. The darkest areas on the left and right of the plots indicate the regions of no current. The bright lines located outside these regions correspond to a fine structure of current steps visible near the voltage thresholds.

This behaviour is the signature of a single-electron transistor¹⁴, a device containing a small island which is attached to electrodes by tunnel barriers and whose charge state can be tuned using a gate voltage. In this case the island is a single Co ion. For most values of V_g , the charge state of the ion is stable at low V (dark regions). An electron does not have sufficient energy to tunnel onto the island and therefore current is blocked (Coulomb blockade). The bright lines that define the boundaries of the Coulomb-blockade regions illustrate the tunnelling thresholds for transitions between charge states. Conductance in the vicinity of $V = 0$ is allowed at a value of

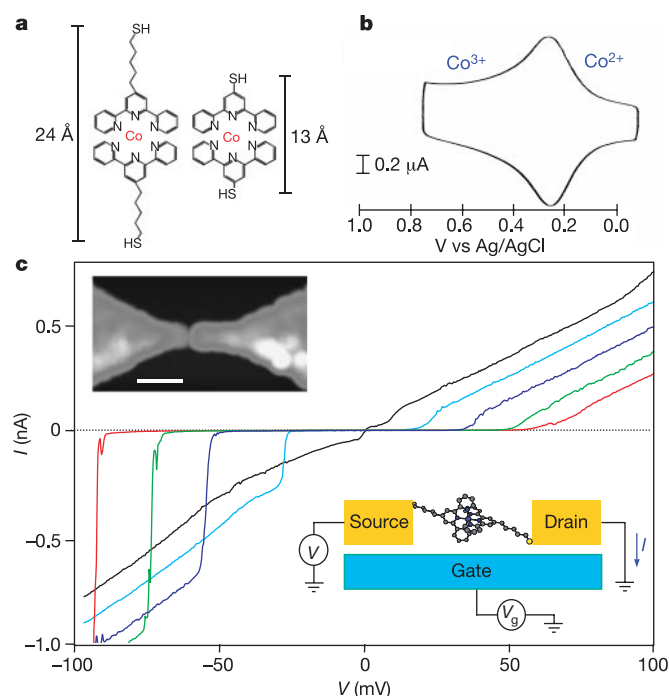


Figure 1 The molecules used in this study and their electronic properties. **a**, Structure of $[\text{Co}(\text{tpy}-(\text{CH}_2)_5\text{-SH})_2]^{2+}$ (where $\text{tpy}-(\text{CH}_2)_5\text{-SH}$ is 4'-(5-mercaptopentyl)-2,2':6',2''-terpyridinyl) and $[\text{Co}(\text{tpy-SH})_2]^{2+}$ (where tpy-SH is 4'-(mercapto)-2,2':6',2''-terpyridinyl). The scale bars show the lengths of the molecules as calculated by energy minimization. **b**, Cyclic voltammogram of $[\text{Co}(\text{tpy-SH})_2]^{2+}$ in 0.1 M tetra-*n*-butylammonium hexafluorophosphate/acetonitrile showing the $\text{Co}^{2+}/\text{Co}^{3+}$ redox peak. **c**, I - V curves of a $[\text{Co}(\text{tpy}-(\text{CH}_2)_5\text{-SH})_2]^{2+}$ single-electron transistor at different gate voltages (V_g) from -0.4 V (red) to -1.0 V (black) with $\Delta V_g \approx -0.15$ V. Upper inset, a topographic atomic force microscope image of the electrodes with a gap (scale bar, 100 nm). Lower inset, a schematic diagram of the device.

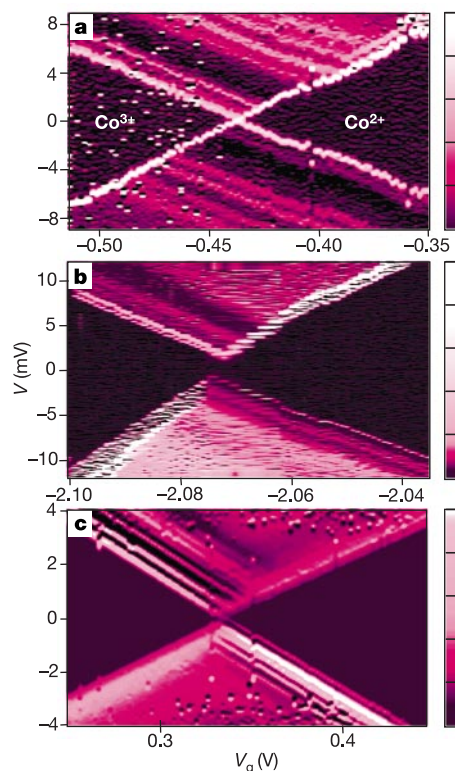


Figure 2 Colour-scale plots of differential conductance ($\partial I/\partial V$) as a function of the bias voltage (V) and the gate voltage (V_g) for three different $[\text{Co}(\text{tpy}-(\text{CH}_2)_5\text{-SH})_2]$ single-electron transistors at zero magnetic field. Black represents zero conductance and white the maximum conductance. The maxima of the scales are 5 nS in **a**, 10 nS in **b**, and 500 nS in **c**. The $\partial I/\partial V$ values were acquired by numerically differentiating individual I - V curves.

gate voltage V_g where the charge states are degenerate. We label the charge states as Co^{2+} and Co^{3+} , in analogy with the electrochemical measurements, and this is supported by a spin analysis presented below.

In control experiments, this behaviour has not been observed for any of 100 bare gold wires or 50 gold wires coated with tpy- $(\text{CH}_2)_5$ -SH linker molecules alone without Co ions. This provides strong evidence that the island of the single-electron transistor is indeed the Co ion. We can be confident that the current near each degeneracy point is due to a single molecule because the degeneracy voltage V_c is different for each molecule owing to local variations in the electrostatic environment. The non-blockaded resistance of devices range from 100 M Ω to ~ 1 G Ω . This is comparable to the resistance recently measured for alkanedithiol molecules whose length is comparable to the linker molecule used here⁷. These results clearly illustrate that the properties of the molecule are reflected in the electrical properties of the single-electron transistor.

Additional lines in Fig. 2 running parallel to the tunnelling thresholds indicate the contributions of excited states to the tunnelling current. Lines that end in the Co^{3+} (Co^{2+}) blockade region correspond to excited levels of the Co^{3+} (Co^{2+}) charge state. The pattern of excited states is qualitatively, but not quantitatively, similar from molecule to molecule. Typically, we observe several lines at energies below 6 meV. No additional lines are resolved between about 6 and 30 meV, at which point additional strong peaks are seen.

A notable feature of the excited-state spectra is that the pattern of low-lying excitations is the same for both charge states of a given molecule. This, together with the small energy scale, suggests that the low-energy excitations are not associated with different electronic configurations. In order to test whether the excitations may be associated with the emission of a phonon⁹, we have calculated the normal modes of the molecule using a quantum chemistry package (HyperChem 7.0). The simulations show normal modes with energies beginning at approximately 1 meV, with a density of ~ 2 modes per meV, in reasonable consistency with our observations. We are currently studying how the vibrational energies depend on the details of electrode attachment in order to address the differences in the excited-state spectra.

We have applied a magnetic field H to determine the magnetic state of the Co ion. Figure 3a shows a colour plot of $\partial I/\partial V$ at a magnetic field of 6 T for the same device as in Fig. 2a. A new excited Co^{2+} level, denoted by a triangle, has split from the Co^{3+} to Co^{2+} ground-state transition. The energy difference between these two states is linear in H , with a slope corresponding to a g -factor of 2.1 ± 0.2 (Fig. 3b). There is no corresponding Co^{3+} excited-state splitting from the Co^{2+} to Co^{3+} ground-state transition.

These results indicate that the Co^{2+} state is spin-degenerate,

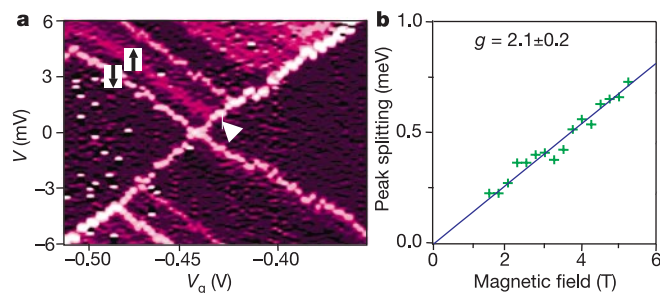


Figure 3 Magnetic-field dependence of the tunnelling spectrum of a $[\text{Co}(\text{tpy}-(\text{CH}_2)_5\text{-SH})_2]$ single-electron transistor. **a**, Differential conductance plot of the device shown in Fig. 2a at a magnetic field of 6 T. There is an extra level (indicated with the triangle) owing to the Zeeman splitting of the lowest energy level of Co^{2+} . The arrows denote the spin of the tunnelling electron. **b**, Magnitude of the Zeeman splitting as a function of magnetic field.

whereas the Co^{3+} state is not. An unambiguous identification of the Co^{2+} ground state as having a total spin quantum number of $S = 1/2$ and the Co^{3+} ground state as $S = 0$ is indicated by an analysis of the tunnelling current amplitudes. The lower-energy Zeeman-split state in Fig. 3a carries a current of 1.0 pA, and the second is nearly equal, 0.8 pA. Nearly equal currents are expected for $S = 0$ to $S = 1/2$ tunnelling, for a tunnelling threshold across the higher-resistance tunnel junction^{15,16}. For any higher spin, the current carried by the second state would be suppressed by a Clebsch–Gordan coefficient by at least a factor of 2 compared to the first state¹⁷.

The electronic structure inferred above is consistent with the expected electronic structure of the Co ion if its angular momentum is quenched owing to the binding to ligand molecules. A Co^{2+} ion ($3d^7$) has an odd number of electrons and possesses Kramers-degenerate states that will split in a magnetic field, whereas Co^{3+} ($3d^6$) has an even number of electrons and may have a total spin $S = 0$ so that it will not undergo Zeeman splitting¹⁸. Previous magnetization studies in bulk material suggest that Co^{2+} in the molecule is $S = 1/2$ at cryogenic temperatures¹⁹. Our measurements provide a confirmation of this result.

We now turn to the results for the shorter molecule, $[\text{Co}(\text{tpy-SH})_2]$, where we expect significantly larger conductances owing to the shorter tether length. In fact, we observe conductances that are large enough to enable us to see directly when a molecule becomes inserted in the gap between the electrodes (Fig. 4a). During the course of electromigration, the conductance initially decreases below the conductance quantum ($2e^2/h$), indicating a tunnelling gap between the electrodes. If the voltage is increased further, the current often suddenly increases by up to a factor of 10 (red dot, Fig. 4a). This behaviour is not observed for bare gold electrodes. We

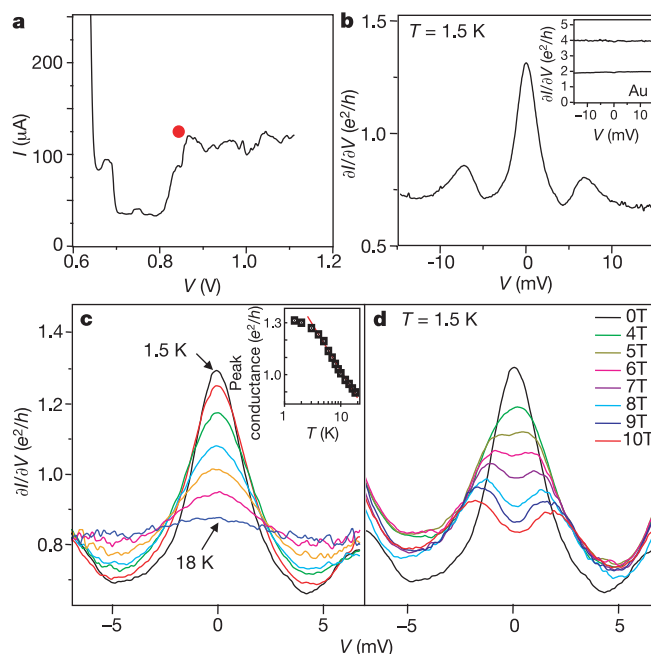


Figure 4 Devices made using the shorter molecule, $[\text{Co}(\text{tpy-SH})_2]^{2+}$, exhibit the Kondo effect. **a**, Breaking trace of a gold wire with adsorbed $[\text{Co}(\text{tpy-SH})_2]^{2+}$ at 1.5 K. After the wire is broken the current level suddenly increases (red dot) owing to the incorporation of a molecule in the gap. This is not seen for bare gold wires. **b**, Differential conductance of a $[\text{Co}(\text{tpy-SH})_2]^{2+}$ device at 1.5 K showing a Kondo peak. The inset shows $\partial I/\partial V$ plots for bare gold point contacts for comparison. **c**, The temperature dependence of the Kondo peak for the device shown in **b**. The inset shows the $V = 0$ conductance as a function of temperature. The peak height decreases approximately logarithmically with temperature and vanishes around 20 K. **d**, Magnetic-field dependence of the Kondo peak. The peak splitting varies linearly with magnetic field.

therefore interpret the jump as the inclusion of at least one molecule in the gap between electrodes²⁰. We stop the electromigration process once this happens, and study the devices at lower V .

The differential conductance $\partial I/\partial V$ for one such device is shown in Fig 4b. The most notable property is a peak in $\partial I/\partial V$ at $V = 0$. The peak has a logarithmic temperature dependence between 3 and 20 K (Fig. 4c). The peak also splits in an applied magnetic field (Fig. 4d), with a splitting equal to $2g\mu_B H$, where $g \approx 2$ and μ_B is the Bohr magneton. This peak has been clearly observed in about 30% of ~ 100 wires broken with the molecule, but is absent in clean gold electrodes, showing that it arises from the presence of the molecule.

The logarithmic temperature dependence and the magnetic-field splitting indicate that the peak is due to the Kondo effect²¹. The Kondo effect is the formation of a bound state between a local spin on an island and the conduction electrons in the electrodes that enhances the conductance at low biases. The observation of the Kondo effect is consistent with the identification of $S = 1/2$ for the Co^{2+} ion given above. By setting the low-temperature full-width at half-maximum of the Kondo peak equal to $2k_B T_K/e$, where T_K is the Kondo temperature^{22,23}, we estimate that T_K in different devices varies between 10 and 25 K. These large Kondo temperatures indicate that the coupling between the localized state and the island is strong, consistent with the high conductances found for the shorter linker molecule. In three devices the gate coupling was strong enough that T_K was increased by sweeping V_g to more negative values, indicating that the energy of the electronic state on the ion is tuned closer to the Fermi level. These devices thus provide an atomic-scale realization of the Kondo model, with Kondo temperatures higher than have been previously reported in quantum dots²⁴ or nanotubes²³.

We have made transistors from a single molecular complex in which one cobalt ion is connected to gold electrodes by organic barriers. By tuning the length of the organic barrier, we are able to control the coupling between the ion and the electrodes. For relatively long linker molecules, giving weak coupling, the molecule functions as a single-electron transistor. For stronger coupling, we observe Kondo-assisted tunnelling. We believe that the ability to design the electronic states of a molecular device using chemical techniques, together with the ability to measure individual molecules, will be important in molecular electronics and in the study of the physics of nanometre-scale systems. □

Methods

Synthesis of the molecules

The longer molecule investigated, $[\text{Co}(\text{tpy}-(\text{CH}_2)_5\text{-SH})_2]^{2+}$, was synthesized from an ethanolic solution of 4'-(5-mercaptopentyl)-2,2':6',2''-terpyridinyl ($\text{tpy}-(\text{CH}_2)_5\text{-SH}$) and aqueous CoCl_2 (ref. 12). The shorter molecule, $[\text{Co}(\text{tpy-SH})_2]^{2+}$, was a complex of cobalt with 4'-(mercapto)-2,2':6',2''-terpyridinyl (tpy-SH). The tpy-SH ligand was prepared from 4'-chloro-2,2':6',2''-terpyridinyl and sodium ethanethiolate by a nucleophilic aromatic substitution followed by nucleophilic aliphatic substitution to give the thiolate anion and subsequent protonation to give the desired compound^{25,26}.

Received 18 February; accepted 29 April 2002; doi:10.1038/nature00791.

- Aviram, A. & Ratner, M. A. Molecular rectifiers. *Chem. Phys. Lett.* **29**, 277–283 (1974).
- Chen, J., Reed, M. A., Rawlett, A. M. & Tour, J. M. Large on-off ratios and negative differential resistance in a molecular electronic device. *Science* **286**, 1550–1552 (1999).
- Collier, C. P. *et al.* Electronically configurable molecular-based logic gates. *Science* **285**, 391–394 (1999).
- Reed, M. A., Zhou, C., Muller, C. J., Burgin, T. P. & Tour, J. M. Conductance of a molecular junction. *Science* **278**, 252–254 (1997).
- Kerker, C. *et al.* Electron transport through a metal-molecule-metal junction. *Phys. Rev. B* **59**, 12505–12513 (1999).
- Bumm, L. A. *et al.* Are single molecular wires conducting? *Science* **271**, 1705–1707 (1996).
- Cui, X. D. *et al.* Reproducible measurement of single-molecule conductivity. *Science* **294**, 571–574 (2001).
- Dekker, C. Carbon nanotubes as molecular quantum wires. *Phys. Today* **52**, 22–28 (1999).
- Park, H. *et al.* Nanomechanical oscillations in a single- C_{60} transistor. *Nature* **407**, 57–60 (2000).
- Schön, J. H., Meng, H. & Bao, Z. Field-effect modulation of the conductance of single molecules. *Science* **294**, 2138–2140 (2001).
- Bard, A. J. & Faulkner, L. R. *Electrochemical Methods: Fundamentals and Applications* (Wiley & Sons, New York, 2001).
- Maskus, M. & Abruna, H. D. Synthesis and characterization of redox-active metal complexes

- sequentially self-assembled onto gold electrodes via a new thiol-terpyridine ligand. *Langmuir* **12**, 4455–4462 (1996).
- Park, H., Lim, A. K. L., Park, J., Alivisatos, A. P. & McEuen, P. L. Fabrication of metallic electrodes with nanometer separation by electromigration. *Appl. Phys. Lett.* **75**, 301–303 (1999).
- Grabert, H. & Devoret, M. H. *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures* (Plenum, New York, 1992).
- Deshmukh, M. M., Bonet, E., Pasupathy, A. N. & Ralph, D. C. Equilibrium and nonequilibrium electron tunneling via discrete quantum states. *Phys. Rev. B* **65**, 073301-1–073301-4 (2002).
- Bonet, E., Deshmukh, M. M. & Ralph, D. C. Solving rate equations for electron tunneling via discrete quantum states. *Phys. Rev. B* **65**, 045317-1–045317-10 (2002).
- Akera, H. Coulomb staircase and total spin in quantum dots. *Phys. Rev. B* **60**, 10683–10686 (1999).
- Ralph, D. C., Black, C. T. & Tinkham, M. Gate-voltage studies of discrete electronic states in aluminum nanoparticles. *Phys. Rev. Lett.* **78**, 4087–4090 (1997).
- Oshio, H., Spiering, H., Ksenofontov, V., Renz, F. & Guetlich, P. Electronic relaxation phenomena following $^{57}\text{Co}(\text{EC})^{57}\text{Fe}$ nuclear decay in $[\text{Mn}^{\text{II}}(\text{terpy})_2](\text{ClO}_4)_2 \cdot 1/2\text{H}_2\text{O}$ and in the spin crossover complexes $[\text{Co}^{\text{II}}(\text{terpy})_2]\text{X}_2 \cdot n\text{H}_2\text{O}$ ($\text{X} = \text{Cl}$ and ClO_4): A Moessbauer emission spectroscopic study. *Inorg. Chem.* **40**, 1143–1150 (2001).
- Bezryadin, A., Dekker, C. & Schmid, G. Electrostatic trapping of single conducting nanoparticles between nanoelectrodes. *Appl. Phys. Lett.* **71**, 1273–1275 (1999).
- Wolf, E. L. *Principles of Electron Tunneling Spectroscopy* Ch. 8 (Oxford Univ. Press, Oxford, 1989).
- van der Wiel, W. G. *et al.* The Kondo effect in the unitary limit. *Science* **289**, 2105–2108 (2000).
- Nygård, J., Cobden, D. H. & Lindelof, P. E. Kondo physics in carbon nanotubes. *Nature* **408**, 342–346 (2000).
- Goldhaber-Gordon, D. *et al.* Kondo effect in a single-electron transistor. *Nature* **391**, 156–159 (1998).
- Testaferri, L., Tiecco, M., Tingoli, M., Chianelli, D. & Montanucci, M. Simple syntheses of aryl alkyl thioethers and of aromatic thiols from unactivated aryl halides and efficient methods for selective dealkylation of aryl alkyl ethers and thioethers. *Synthesis (Stuttgart)* **9**, 751–755 (1983).
- Mathis, J. M. & Pallenberg, A. J. Preparation of novel, functionalized 1,10-phenanthrolines. *Synth. Commun.* **27**, 2943–2951 (1997).

Acknowledgements

We thank E. Smith, M. Brink and J.-Y. Park for help with measurements, and M. Deshmukh for discussions. This work was supported by NSF, through individual-investigator grants, the Cornell Center for Materials Research, and the use of the National Nanofabrication Users Network. Support was also provided by the Packard Foundation, the US Department of Energy and Department of Education GAANN fellowships.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.M. (e-mail: mceuen@ccmr.cornell.edu) or D.R. (e-mail: ralph@ccmr.cornell.edu).

Kondo resonance in a single-molecule transistor

Wenjie Liang*, Matthew P. Shores†, Marc Bockrath*, Jeffrey R. Long† & Hongkun Park*

* Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

† Department of Chemistry, University of California, Berkeley, California 94720, USA

When an individual molecule¹, nanocrystal^{2–4}, nanotube^{5,6} or lithographically defined quantum dot⁷ is attached to metallic electrodes via tunnel barriers, electron transport is dominated by single-electron charging and energy-level quantization⁸. As the coupling to the electrodes increases, higher-order tunnelling and correlated electron motion give rise to new phenomena^{9–19}, including the Kondo resonance^{10–16}. To date, all of the studies of Kondo phenomena in quantum dots have been performed on systems where precise control over the spin degrees of freedom is difficult. Molecules incorporating transition-metal atoms provide powerful new systems in this regard, because the spin and orbital degrees of freedom can be controlled through well-defined chemistry^{20,21}. Here we report the observation of the Kondo effect in single-molecule transistors, where an individual divanadium molecule²⁰ serves as a spin impurity. We find that the

Kondo resonance can be tuned reversibly using the gate voltage to alter the charge and spin state of the molecule. The resonance persists at temperatures up to 30 K and when the energy separation between the molecular state and the Fermi level of the metal exceeds 100 meV.

We prepared devices by an extension of the methods previously used in constructing single-C₆₀ (ref. 1) and single-nanocrystal transistors³. Using electron-beam lithography, a narrow gold bridge was fabricated on an aluminium pad with a ~3-nm oxide layer serving as a gate electrode²². The electromigration-induced break-junction technique^{1,3} was then used to create two closely spaced gold electrodes (Fig. 1). Scanning electron microscope imaging and tunnel current measurements reveal that the narrowest gap between the two electrodes is consistently ~1 nm.

Single-molecule transistors containing individual divanadium (V₂) molecules ([*(N,N',N''-trimethyl-1,4,7-triazacyclononane)*₂-V₂(CN)₄(μ-C₄N₄)]: Fig. 1)²⁰ were prepared by depositing a dilute methanol solution of the V₂ molecule onto the gold bridge. To minimize the probability of multiple V₂ molecules bridging two electrodes, we controlled the coverage of molecules on an array of gold electrodes such that <20% of the junctions—7 in an array of 42 broken junctions—showed current-voltage (*I*-*V*) characteristics different from a simple tunnel junction.

Figure 2 shows plots of differential conductance ($\partial I/\partial V$) as a function of bias voltage (*V*) and gate voltage (*V_g*) for two single-V₂

transistors, designated D1 and D2. Two distinct characteristics, which are shared by all seven single-V₂ transistors studied to date, are evident in the behaviour of both devices. Each displays two conductance-gap regions, I and II, bounded by two broad $\partial I/\partial V$ peaks that slope linearly as a function of *V_g*. These peaks cross at *V_g* = *V_c*, at which point the conductance gaps vanish. Moving away from this point, the gaps in both regions continue to widen even beyond *V* = 100 mV. Most significantly, the devices also exhibit a sharp zero-bias $\partial I/\partial V$ peak in region I, whereas this peak is clearly absent in region II. The zero-bias $\partial I/\partial V$ peak was generally found to persist up to *T* ≥ 10 K. Other features observable in D2, such as the enhanced conductance within region I and the $\partial I/\partial V$ peaks outside the conductance gaps, were not shared by all devices (see below).

The observation of conductance gaps that vary linearly with *V_g* provides strong experimental evidence that an individual molecule is responsible for the device behaviour. A conductance gap arises from the finite energy required to add (remove) an electron to (from) the molecule, which is a consequence of single-electron charging and the quantized molecular level spacing. The conductance gap changes linearly and reversibly as a function of *V_g* because a more positive *V_g* stabilizes an additional electron on the V₂ molecule. At the charge degeneracy point, *V_g* = *V_c*, where the total energy is the same for two different charge states of the molecule, the conductance gap disappears. As *V_g* traverses *V_c* in the positive direction, the equilibrium number of charges on the molecule increases by one electron.

Certainly the most prominent feature of the data is the appear-

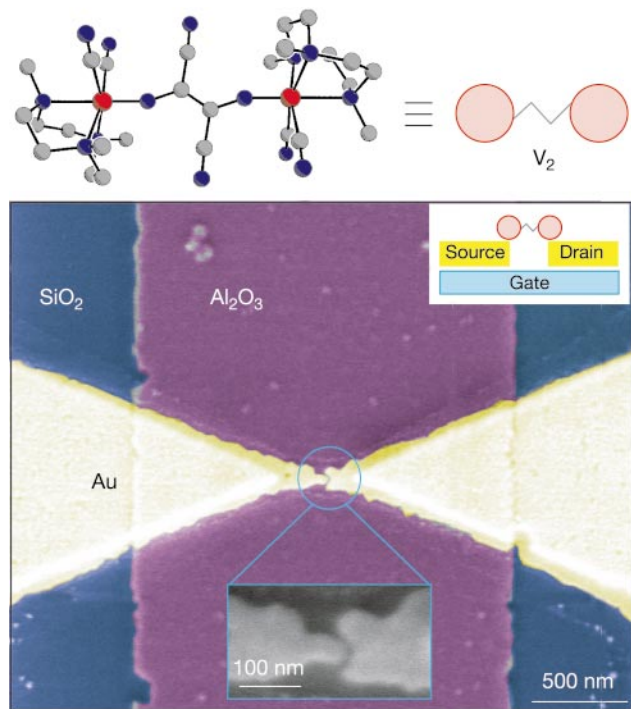


Figure 1 Fabrication of single-molecule transistors incorporating individual divanadium molecules. Top left, the structure of [*(N,N',N''-trimethyl-1,4,7-triazacyclononane)*₂-V₂(CN)₄(μ-C₄N₄)] (the V₂ molecule) as determined by X-ray crystallography; red, grey and blue spheres represent respectively V, C and N atoms. Top right, the schematic representation of this molecule. Main panel, scanning electron microscope image (false colour) of the metallic electrodes fabricated by electron beam lithography and the electromigration-induced break-junction technique. The image shows two gold electrodes separated by ~1 nm above an aluminium pad, which is covered with an ~3-nm-thick layer of aluminium oxide. The whole structure was defined on a silicon wafer. The bright yellow regions correspond to a gold bridge with a thickness of 15 nm and a minimum lateral size of ~100 nm. The paler yellow regions represent portions of the gold electrodes with a thickness of ~100 nm. Main panel inset, schematic diagram of a single-V₂ transistor.

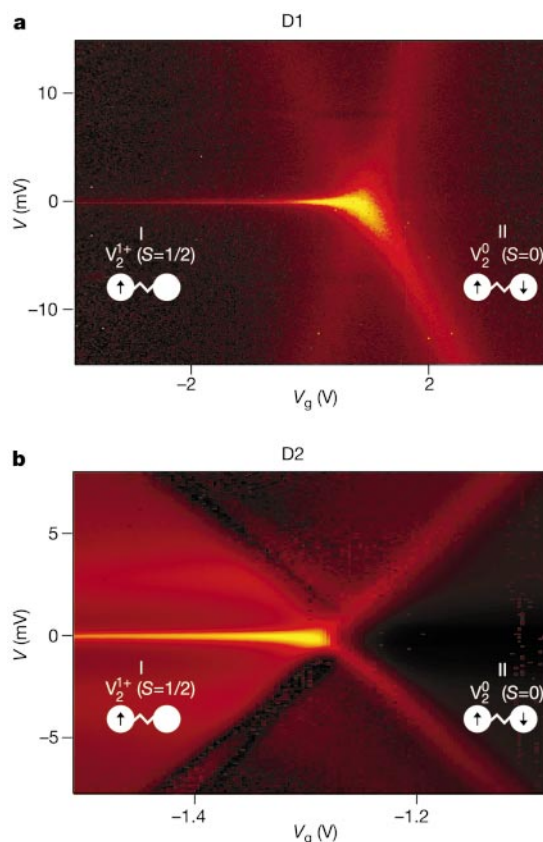


Figure 2 Plots of differential conductance ($\partial I/\partial V$) as a function of bias voltage (*V*) and gate voltage (*V_g*) obtained from two different single-V₂ transistors D1 (**a**) and D2 (**b**). Both measurements were performed at *T* = 300 mK. The $\partial I/\partial V$ values are represented by the colour scale, which changes in **a**, from dark red (0) to bright yellow ($1.55 e^2/h$) and in **b**, from dark red (0) to bright yellow ($1.3 e^2/h$). The value of e^2/h is $38.8 \mu S$ or $(25.8 k\Omega)^{-1}$. The labels I and II mark two conductance-gap regions, and the diagrams indicate the charge and spin states of the V₂ molecule in each region.

ance of the sharp zero-bias $\partial I/\partial V$ peak in region I. This feature is strongly reminiscent of the Kondo resonance observed previously in lithographically defined quantum dots^{10–14} and single-walled carbon nanotubes^{15,16}. The Kondo resonance is a many-electron phenomenon resulting from the exchange interaction between a localized spin and the conduction electrons in metallic electrodes^{17–19,23,24}, and it appears only when the electronic state of a quantum dot has non-zero spin (S) or degeneracy.

One signature of a Kondo resonance arising from non-zero S is the splitting of its peak position in an applied magnetic field, B , by twice the ordinary Zeeman splitting¹⁹. Figure 3a depicts a plot of $\partial I/\partial V$ versus V and B , and clearly shows that the zero-bias $\partial I/\partial V$ peak in region I splits in V under a magnetic field. The magnitude of this splitting is $230 \mu\text{V T}^{-1}$ or $2g\mu_B B/e$ (where $g = 2$ is the g -factor of the molecule²⁰, and μ_B is the Bohr magneton), consistent with the expected splitting of a Kondo resonance. As the temperature is raised, the peak height of the resonance is expected to decrease in a characteristic fashion while its peak width increases (see below for further discussion)^{12,17,24}. Such behaviour is evident in the data presented in Fig. 4. These observations therefore firmly establish that the zero-bias $\partial I/\partial V$ peaks in region I arise from a Kondo resonance in the single- V_2 transistors.

The absence of a Kondo resonance in region II suggests, on the other hand, that the requisite spin degeneracy is lost when the charge state of the V_2 molecule changes by the addition of one electron. Insight into the origin of this behaviour can be obtained by inspecting Fig. 3b, which displays a $\partial I/\partial V$ – V_g plot analogous to those in Fig. 2, but with data measured at $B = 8$ T. In particular, Fig. 3b shows that the $\partial I/\partial V$ peaks defining region II exhibit an ordinary Zeeman splitting of $g\mu_B B/e = 115 \mu\text{V T}^{-1}$, whereas the peaks defining region I show no such splitting. This splitting pattern is characteristic of the ground-state spin change of $\Delta S = -1/2$ upon the addition of one electron²⁵. Comparison of Figs 2b and 3b further shows that the charge degeneracy point ($V_g = V_c$) moves in the positive V_g direction as the magnetic field is increased, signifying that the ground state of the molecule in region I is stabilized relative to that in region II. Analysis of this shift on the basis of the known electrostatic coupling to the gate electrode indicates that the energy stabilization amounts to $\sim 100 \mu\text{V T}^{-1}$, again consistent with a reduction of the spin by $1/2$.

In the neutral V_2 molecule, both vanadium atoms exist in a $4+$ oxidation state, each possessing one valence d electron²⁰ (Fig. 3c). Electrochemical studies show that in organic solvents three oxidation states of the V_2 molecule are accessible²⁰: neutral V_2^0 , positive V_2^+ , and negative V_2^- . Magnetic measurements performed on bulk samples indicate that the ground state of V_2^0 is a spin-singlet ($S = 0$) owing to the antiferromagnetic coupling between vanadium centres, while that of V_2^- is a spin-quadruplet ($S = 3/2$) owing to the resonant exchange of the added electron²⁰. Although the V_2^+ ion has not yet been isolated in bulk, the ground state of V_2^+ is expected to be a spin-doublet ($S = 1/2$) as it possesses only one d electron. These results, combined with the information that the ground-state spin value reduces by $1/2$ upon one-electron addition, indicate that the V_2^+ and V_2^0 species are responsible for the device behaviour in regions I and II, respectively, as shown in Fig. 2.

Determination of the spin multiplicity permits a quantitative analysis of the temperature dependence of the Kondo effect (Fig. 4). Previous theoretical^{17–19,23,24} and experimental^{10–16} studies have shown that the Kondo physics of a non-degenerate spin- $1/2$ system is characterized by a single temperature scale known as the Kondo temperature (T_K). In the limit of a large charging energy U ($U > 1$ eV in the V_2 molecule) and $\varepsilon \gg \Gamma$, T_K is expected to follow the functional form^{7,26} $T_K = 0.5(TU)^{1/2} \exp(-\pi\varepsilon/\Gamma)$. Here Γ is the level width due to the tunnel coupling to metallic electrodes, and $-\varepsilon$ is the energy of the localized electron measured relative to the metal Fermi level. Although this expression is derived in the Kondo regime ($\varepsilon/\Gamma \gg 1$), previous experimental studies have shown that the

expression also adequately describes data in the crossover region between the mixed-valence ($\varepsilon/\Gamma \leq 0.5$) and the Kondo regimes^{12,14}. The temperature dependence of the Kondo peak height (G_K) follows an approximate scaling form given by $G_K(T) = G_0/(1 + (2^{1/s} - 1)T^2/T_K^2)^s$, where G_0 is a constant¹². The value of s depends on the value of ε/Γ , and it attains its asymptotic value of 0.22 when $\varepsilon/\Gamma \gg 1$. At low temperatures, the width of the Kondo peak is expected to saturate at $\sim k_B T_K/e$, where k_B is the Boltzmann constant¹¹.

T_K and the resonance width extracted from device D3 are presented as a function of ε/Γ in Fig. 4b. The value of T_K was determined by fitting the measured G_K – T curves to the above scaling expression using G_0 and s as fit parameters, and the peak width was measured at the full-width at half-maximum. We note that the T_K value close to $V_g = V_c$ exceeds 30 K, representing (to our knowledge) the highest Kondo temperature reported to date for a quantum-dot type system. Far away from $V_g = V_c$, the peak widths converge to an asymptotic value owing to thermal broadening. The Kondo resonance nevertheless persists even when ε exceeds 100 meV ($\varepsilon/\Gamma > 3$).

Figure 4b shows that, as ε is tuned away from the Fermi level, both T_K and the peak width decay as expected, but they do not follow a single-exponential form. The values of ε/Γ range from 0.1 to 2, thus spanning the mixed-valence regime ($\varepsilon/\Gamma \leq 0.5$) and crossing over to the Kondo regime ($\varepsilon/\Gamma \gg 1$) (refs 12, 24, 26, 27). As shown by two lines in Fig. 4b, the decay can be approximated by $\exp(-3\varepsilon/\Gamma)$ for $0 < \varepsilon/\Gamma < 0.5$ and then rolls over to $\exp(-1.3\varepsilon/\Gamma)$

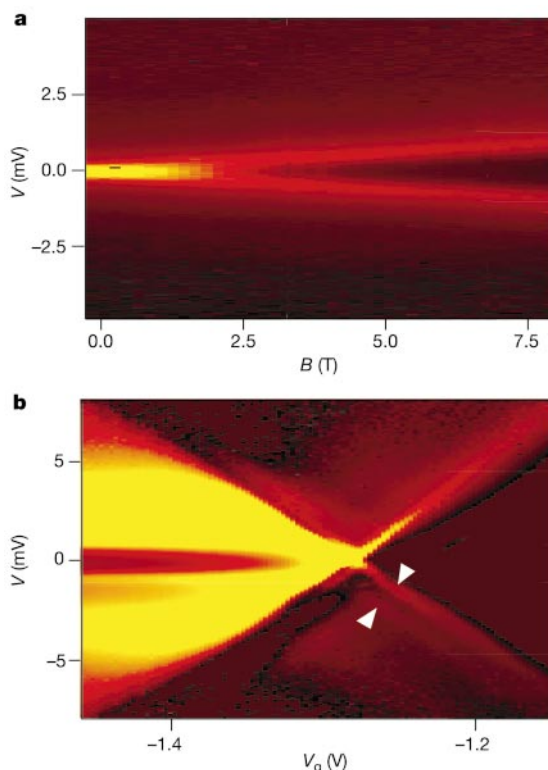


Figure 3 Transport data obtained from single- V_2 transistors in an applied magnetic field (B). **a**, A $\partial I/\partial V$ plot as a function of V and B obtained from D1 at $V_g = -0.1$ V and at $T = 300$ mK. The $\partial I/\partial V$ values are represented by a colour scale that varies from dark red (0) to bright yellow ($1.3 e^2/h$). **b**, A $\partial I/\partial V$ plot as a function of V and V_g obtained from D2 at $B = 8$ T and at $T = 300$ mK. White arrows indicate the two $\partial I/\partial V$ peaks that arise from a Zeeman splitting. To clearly illustrate weak Zeeman-split features, the colour scale has been changed from that in Fig. 2a and varies from dark red (0) to bright yellow ($0.55 e^2/h$).

for $0.5 < \varepsilon/\Gamma < 2$. In the range of $0.5 < \varepsilon/\Gamma < 2$, the decay is approximately a factor of two slower than what has been observed previously for a non-degenerate spin-1/2 system^{12,14}.

One possible explanation for this behaviour is that not only the spin but also orbital degrees of freedom may be required to explain the Kondo physics observed in single- V_2 transistors. As discussed previously, the V_2^+ ion contains two equivalent vanadium sites and hence it can provide doubly degenerate orbitals for a single d electron. Such degeneracy is known to be important from studies of lanthanide impurities in metal hosts^{23,28,29}, and theoretical calculations predict that the Kondo temperature scales as $T_K \propto \exp(-\pi\varepsilon/2\Gamma)$ for doubly degenerate orbital cases^{23,28,29}, therefore accounting for our experimental data in a satisfactory fashion. Nevertheless, given the experimental and theoretical uncertainties, this conclusion remains speculative, and further investigation is necessary to resolve the role of orbital degeneracy in the single- V_2 -transistor system.

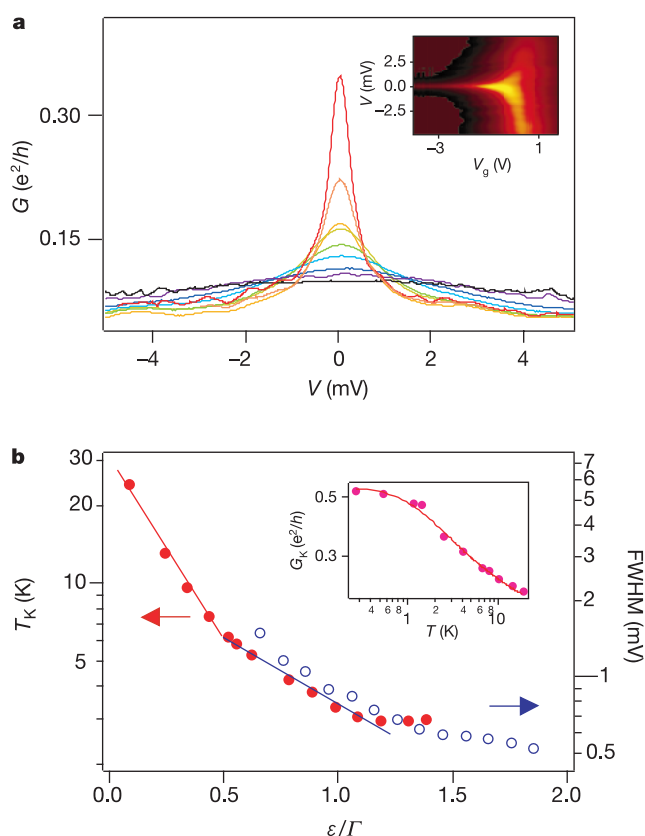


Figure 4 Temperature-dependent transport data from device D3. **a**, A plot of conductance (G) versus V with $V_g = -2.25$ V at various temperatures. The temperatures of the measurements (in K) are $T = 0.3, 1.0, 2.0, 3.1, 4.2, 6.3, 9.0, 14$ and 20 , in order of decreasing peak height. Inset, a dI/dV plot as a function of V and V_g at $T = 300$ mK. The colour scale changes from dark red (0) to bright yellow ($0.55 e^2/h$). **b**, The Kondo temperature (T_K ; filled red circles) and the Kondo peak width determined by the full-width at half-maximum (open blue circles) plotted against ε/Γ in a logarithmic scale. Here $-\varepsilon$ is the energy of the localized electron measured relative to the Fermi level of the metal, and Γ is the level width due to the tunnel coupling to the metallic electrodes. Measurements of the dI/dV peak widths and slopes that define conductance gaps (inset in **a**) show that Γ is ~ 30 mV and the gate coupling is $\alpha = C_g/C_\Sigma = 30 \text{ meV}V_g^{-1}$ (C_g is the capacitance to the gate, and C_Σ is the total capacitance). This value of α allows the conversion of V_g to ε , because $\varepsilon = \alpha(V_c - V_g)$. We estimate that the values of Γ and ε are accurate to within 20%. Red and blue lines are proportional to $\exp(-3\varepsilon/\Gamma)$ and $\exp(-1.3\varepsilon/\Gamma)$, respectively. As ε/Γ exceeds 2, T_K and the peak widths approach respective asymptotic values. Inset, plot of the Kondo peak height (G_K) as a function of temperature at $\varepsilon = 0.43\Gamma$. The line is a fit to the formula $G_K(T) = G_0/(1 + (2^{1/s} - 1)T^2/T_K^2)^s$, yielding $T_K = 12.2$ K and $s = 0.19$.

Overall, the behaviour of a single- V_2 transistor is well described by the Coulomb blockade model and a Kondo resonance originating from the V_2^+ ion. However, certain transport characteristics of these transistors remain to be explained. As noted previously, device D2 exhibits strongly enhanced conductance values within the nominal conductance-gap region I. The bias position of this feature does not depend on V_g , suggesting that it arises from inelastic cotunnelling events^{16,30}. In addition, some devices exhibit one or more $\partial I/\partial V$ peaks outside the conductance gaps. Both observations suggest that the excited electronic or vibrational states may participate in electron transport¹.

The present study shows that molecules can provide a Kondo system where critical parameters of Kondo physics, such as the spin and orbital degrees of freedom, are defined by chemical synthesis. With the recent advances of synthetic methodology, the preparation of molecular clusters possessing adjustable magnetic properties is becoming feasible²¹. Future investigations of such species are expected to provide detailed insight into electron transport through a molecular system where the spin and orbital degeneracies are precisely controlled. □

Received 18 February; accepted 29 April 2002; doi:10.1038/nature00790.

1. Park, H. *et al.* Nano-mechanical oscillations in a single- C_{60} transistor. *Nature* **407**, 57–60 (2000).
2. Klein, D. L. *et al.* A single-electron transistor made from a cadmium selenide nanocrystal. *Nature* **389**, 699–701 (1997).
3. Park, H. *et al.* Fabrication of metallic electrodes with nanometer separation by electromigration. *Appl. Phys. Lett.* **75**, 301–303 (1999).
4. Banin, U., Cao, Y., Katz, D. & Millo, O. Identification of atomic-like electronic states in indium arsenide nanocrystal quantum dots. *Nature* **400**, 542–544 (1999).
5. Tans, S. J. *et al.* Individual single-wall carbon nanotubes as quantum wires. *Nature* **386**, 474–476 (1997).
6. Bockrath, M. *et al.* Single-electron transport in ropes of carbon nanotubes. *Science* **275**, 1922–1925 (1997).
7. Kouwenhoven, L. P. *et al.* Excitation spectra of circular few-electron quantum dots. *Science* **278**, 1788–1792 (1997).
8. Grabert, H. & Devoret, M. H. *Single Charge Tunneling* (Plenum, New York, 1992).
9. Liang, W. *et al.* Fabry-Perot interference in a nanotube electron waveguide. *Nature* **411**, 665–669 (2001).
10. Goldhaber-Gordon, D. *et al.* Kondo effect in a single-electron transistor. *Nature* **391**, 156–159 (1998).
11. Cronenwett, S. M., Oosterkamp, T. H. & Kouwenhoven, L. P. A tunable Kondo effect in quantum dots. *Science* **281**, 540–544 (1998).
12. Goldhaber-Gordon, D. *et al.* From the Kondo regime to the mixed-valence regime in a single-electron transistor. *Phys. Rev. Lett.* **81**, 5225–5228 (1998).
13. Schmid, J., Weis, J., Eberl, K. & Klitzing, K. v. Absence of odd-even parity behaviour for Kondo resonances in quantum dots. *Phys. Rev. Lett.* **84**, 5824–5827 (2000).
14. van der Wiel, W. G. *et al.* The Kondo effect in the unitary limit. *Science* **289**, 2105–2108 (2000).
15. Nygard, J., Cobden, D. H. & Lindelof, P. E. Kondo physics in carbon nanotubes. *Nature* **408**, 342–346 (2000).
16. Liang, W., Bockrath, M. & Park, H. Shell filling and exchange coupling in metallic single-walled carbon nanotubes. *Phys. Rev. Lett.* **88**, 126801–1–126801–4 (2002).
17. Glazman, L. I. & Raikh, M. E. Resonant Kondo transparency of a barrier with quasilocal impurity states. *JETP Lett.* **47**, 452–455 (1988).
18. Ng, T. K. & Lee, P. A. On-site Coulomb repulsion and resonant tunneling. *Phys. Rev. Lett.* **61**, 1768–1771 (1988).
19. Meir, Y., Wingreen, N. S. & Lee, P. A. Low-temperature transport through a quantum dot: The Anderson model out of equilibrium. *Phys. Rev. Lett.* **70**, 2601–2604 (1993).
20. Shores, M. P. & Long, J. R. Tetracyanide-bridged vanadium complexes: Redox switching between strong antiferromagnetic and strong ferromagnetic coupling. *J. Am. Chem. Soc.* **124**, 3512–3513 (2002).
21. Shores, M. P., Sokol, J. J. & Long, J. R. Nickel(II)-molybdenum(III)-cyanide clusters: Synthesis and magnetic behaviour of species incorporating $[(Me_3tacn)Mo(CN)_3]$. *J. Am. Chem. Soc.* **124**, 2279–2292 (2002).
22. Bachtold, A., Hadley, P., Nakanishi, T. & Dekker, C. Logic circuits with carbon nanotube transistors. *Science* **294**, 1317–1320 (2001).
23. Schlottmann, P. Some exact results for dilute mixed-valent and heavy-fermion systems. *Phys. Rep.* **181**, 1–119 (1989).
24. Hewson, A. C. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, Cambridge, 1993).
25. Cobden, D. H. *et al.* Spin splitting and even-odd effects in carbon nanotubes. *Phys. Rev. Lett.* **81**, 681–684 (1998).
26. Haldane, F. D. M. Scaling theory of the asymmetric Anderson model. *Phys. Rev. Lett.* **41**, 416–419 (1978).
27. Wingreen, N. S. & Meir, Y. Anderson model out of equilibrium: Noncrossing-approximation approach to transport through a quantum dot. *Phys. Rev. B* **49**, 11040–11052 (1994).
28. Lin, H. Q. & Hirsch, J. E. Magnetic properties of a degenerate Anderson impurity. *Phys. Rev. B* **37**, 1864–1873 (1988).
29. Bonca, J. & Gubernatis, J. E. Quantum Monte Carlo simulations of the degenerate single-impurity Anderson model. *Phys. Rev. B* **47**, 13137–13146 (1993).
30. De Franceschi, S. *et al.* Electron cotunneling in a semiconductor quantum dot. *Phys. Rev. Lett.* **86**, 878–881 (2001).

Acknowledgements

We thank C. Lieber, B. Halperin and D. R. Reichman for discussions. This work was supported by NSF, DARPA, the Dreyfus Foundation, the Packard Foundation, the Research Corporation, and Harvard University (H.P.) and NSF (J.R.L.). M.B. is partially supported by the Department of Physics, Harvard University.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.P. (e-mail: HPark@chemistry.harvard.edu).

Mycorrhizal weathering of apatite as an important calcium source in base-poor forest ecosystems

Joel D. Blum*, Andrea Klau*, Carmen A. Nezat*, Charles T. Driscoll†, Chris E. Johnson†, Thomas G. Siccama‡, Christopher Eagar§, Timothy J. Fahey|| & Gene E. Likens¶

* Department of Geological Sciences, University of Michigan, 425 East University Avenue, Ann Arbor, Michigan 48109, USA

† Department of Civil and Environmental Engineering, Syracuse University, New York 13244, USA

‡ Yale School of Forestry and Environmental Studies, New Haven, Connecticut 06511, USA

§ Northeastern Research Station, USDA Forest Service, Durham, New Hampshire 03824, USA

|| Department of Natural Resources, Cornell University, Ithaca, New York 14853, USA

¶ Institute of Ecosystem Studies, Millbrook, New York 12545, USA

The depletion of calcium in forest ecosystems of the northeastern USA^{1–3} is thought to be a consequence of acidic deposition and to be at present restricting the recovery of forest and aquatic systems^{4–7} now that acidic deposition itself is declining. This depletion of calcium has been inferred from studies^{1–3} showing that sources of calcium in forest ecosystems—namely, atmospheric deposition and mineral weathering of silicate rocks such as plagioclase, a calcium-sodium silicate—do not match calcium outputs observed in forest streams. It is therefore thought that calcium is being lost from exchangeable and organically bound calcium in forest soils. Here we investigate the sources of calcium in the Hubbard Brook experimental forest, through analysis of calcium and strontium abundances and strontium isotope ratios within various soil, vegetation and hydrological pools. We show that the dissolution of apatite (calcium phosphate) represents a source of calcium that is comparable in size to known inputs from atmospheric sources and silicate weathering. Moreover, apatite-derived calcium was utilized largely by ectomycorrhizal tree species, suggesting that mycorrhizae may weather apatite and absorb the released ions directly, without the ions entering the exchangeable soil pool. Therefore, it seems that apatite weathering can compensate for some of the calcium lost from base-poor ecosystems, and should be considered when estimating soil acidification impacts and calcium cycling.

We studied the biogeochemistry of calcium in a mixed conifer/hardwood forest within a 12-ha watershed (W-1) at the Hubbard Brook experimental forest (HBEF), New Hampshire, USA^{2,7,8}. The main calcium-bearing minerals present in soil parent material are plagioclase, K-feldspar, hornblende, pyroxene and apatite. Sequen-

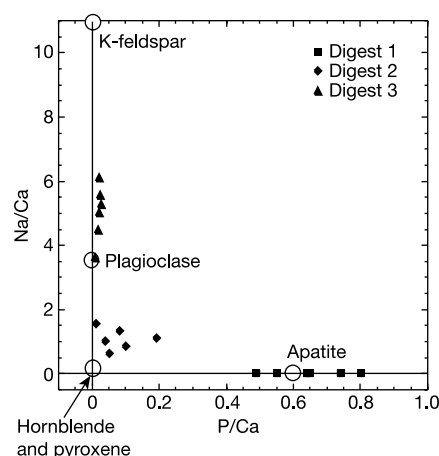


Figure 1 Major-element ratios of sequential digests of six composite C-horizon soil samples and soil minerals. See Methods for details.

tial digestions of soil samples (see Methods) were used to give information on the weathering sources of Ca (Table 1). We first examined the molar ratios of P/Ca versus Na/Ca in C-horizon digestions in comparison with the average composition of the main Ca-bearing minerals (Fig. 1). 'Digest 1' consisted mainly of dissolved apatite, consistent with the much faster ($>10^3$ times) dissolution rate of apatite compared to hornblende, plagioclase and K-feldspar (at pH 2–5)^{9–11}. 'Digest 2' consisted mainly of dissolved hornblende and pyroxene (Fig. 1), and 'digest 3' was mainly dissolved plagioclase and K-feldspar. About 12% of the Ca in the soil parent material occurs in apatite, another 12% is in hornblende and pyroxene, and 76% is in feldspars. The Ca/Sr ratio of digest 1 was very different to that of digests 2 and 3 (Table 1), owing to differences in Sr substitution for Ca in silicate versus phosphate minerals—providing a tracer of apatite-derived Ca. The Ca/Sr ratio of digest 1 for each soil horizon suggests that apatite was depleted from the Oa, E, Bh and Bs1 horizons, but is present in the Bs2 and C horizons (Fig. 2). If we assume that all of the Ca in the C-horizon digest 1 is derived from apatite (Fig. 1), then a Ca/Sr mixing calculation^{12,13} shows that 86% of the Ca released from the Bs2-horizon soils by digest 1 is derived from apatite.

We used Ca/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios to ascertain the original sources of Ca in ecosystem pools^{13–17}. The $^{87}\text{Sr}/^{86}\text{Sr}$ ratio is not changed by ion exchange, plant uptake or transport^{13–19}, and the Ca/Sr ratio has generally been observed to be only slightly changed by these processes (a factor of 1.0–1.6)^{14,18,19}. The C-horizon digest-1 fractions had average Ca/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (respectively 2,580 and 0.72164) that were distinct from the values for the digest-2 (279 and 0.73669) and digest-3 fractions (170 and 0.73352; Fig. 3). This pattern is consistent with our interpretation that digest 1 is dominated by dissolution of apatite, which has characteristically

Table 1 Average exchangeable and sequential digest concentrations

	Ca	K	Mg	Na	Si	P	Ca/Sr*
$\mu\text{mol per g soil}$							
Exchangeable	0.01	0.50	<0.04	<0.04	0.09	0.02	281
Digest 1	24.0	1.45	6.21	0.27	36.9	16.0	2580
Digest 2	24.0	58.1	65.6	25.5	305	1.63	279
Digest 3	151	580	34.7	733	13200	3.04	170
Total soil (sum)	199	640	107	759	13500	20.7	202
Percentage of total soil							
Exchangeable	0.05	0.08	<0.04	<0.01	<0.01	0.11	
Digest 1	12.0	0.23	5.83	0.04	0.27	77.3	
Digest 2	12.0	9.07	61.6	3.36	2.25	7.90	
Digest 3	75.9	90.6	32.5	96.6	97.5	14.7	

Data are shown for six composite C-horizon samples.

*Mole ratio.

high Ca/Sr and low $^{87}\text{Sr}/^{86}\text{Sr}$ (refs 20, 21). In comparison, atmospheric deposition near Hubbard Brook has an average Ca/Sr of 193 and $^{87}\text{Sr}/^{86}\text{Sr}$ of 0.71060 (refs 15, 16). The 'exchangeable' fractions (see Methods) of soil Oa and Bs1 horizons had Ca/Sr ratios similar to atmospheric and soil-digest values, but had $^{87}\text{Sr}/^{86}\text{Sr}$ close to the digest-1 values and intermediate between atmospheric and digest-2 and digest-3 values. Thus the soil cation exchange pool is dominated by a mixture of Ca and Sr from silicate weathering and atmospheric deposition.

In Fig. 3, the foliage samples plot along a line connecting the soil cation exchange pool and apatite, with different species occupying distinct fields. The Ca/Sr ratios were relatively low for sugar maple, intermediate for yellow birch and American beech, and high for balsam fir and red spruce. In contrast, foliar $^{87}\text{Sr}/^{86}\text{Sr}$ was nearly identical for all tree species, and was consistent with both the soil cation exchange pool and apatite weathering values (Fig. 3). Wood ferns, which are shallowly rooted and reflect forest floor values, had Ca/Sr ratios in the same range as maple and birch foliage (Fig. 3). Throughfall from the dominantly spruce/fir areas of the mixed conifer-hardwood canopy is a mixture of atmospheric deposition and spruce/fir foliage, whereas throughfall from the dominantly hardwood areas shows a stronger hardwood foliage influence. The range of soil water Ca/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios for nine lysimeters sampled during each season of 1998, and the average streamwater Ca/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ for weekly samples collected in 1998, are also shown in Fig. 3.

To identify the sources of Ca in various ecosystem pools, we defined the Ca/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of three ecosystem endmembers. For the silicate weathering endmember we used the average value of the soil C-horizon digest 2, because it is the most readily weathered soil silicate fraction. For the apatite weathering endmember, we used the average of the C-horizon digest-1 fractions. The average atmospheric deposition endmember has been measured directly^{15,16}. Apatite weathering does not contribute appreciably to exchangeable Ca in the Bs1 horizon, probably because it has been depleted in the upper horizons of the soil profile (Fig. 2). A mixing calculation for Ca/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ of two endmembers^{12,13} shows that the Bs1 exchange pool contained 55% Ca from silicate weathering and 45% Ca from atmospheric sources (Fig. 3), similar to findings of previous studies^{15,16}.

The composition of the Oa exchange pool, stream water, and vegetation can be explained by a mixture of the Bs1 exchange pool

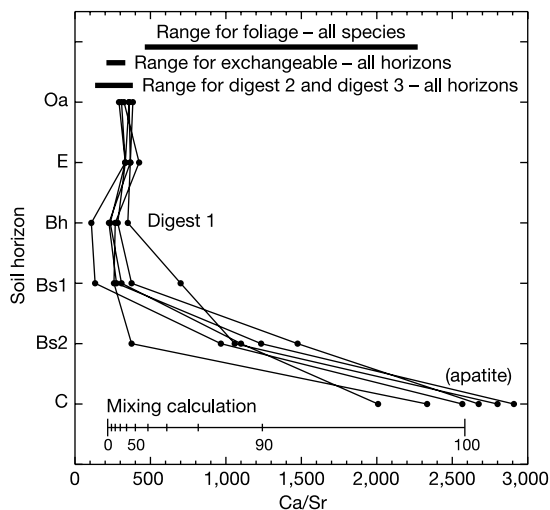


Figure 2 Ca/Sr ratios of the digest 1 for each soil horizon. The bars at the top show the total range of Ca/Sr ratios for all foliage, and for 'exchangeable', 'digest 1', 'digest 2' and 'digest 3' for all horizons (see Methods). The line at the bottom shows the results of a mixing calculation between the Bs1 horizon exchangeable and the average C-horizon digest 1 (that is, apatite) with tick marks for each 10% addition of Ca from apatite.

and apatite (Fig. 3). Mixing calculations provide an estimate of the relative quantity of Ca derived from the Bs1 exchange pool compared to the apatite weathering pool, and show that on average ~95% of the foliar Ca in spruce/fir foliage is derived from apatite, whereas ~80% of beech, ~75% of birch and ~60% of maple and fern Ca is derived from apatite (Figs 2, 3). Also, ~35% of the stream Ca loss from the watershed is estimated to be derived from apatite weathering and ~20% of the exchangeable Ca in the Oa horizon is derived from apatite. Trees at HBEF appear to be accessing Ca and Sr from apatite in the Bs2 horizon, and partially bypassing the soil cation exchange complex. Studies in base-poor northern European coniferous forests have suggested that ectomycorrhizal mycelia are able to penetrate micropores in silicate minerals and take up Ca and P directly from apatite inclusions, while remaining isolated from the soil solution²². This process has only been observed in conifer forests^{22,23} and pot experiments^{23,24}, but may explain the high abundance of apatite-derived Ca in the foliage of the mixed conifer/hardwood forest at HBEF. The shallower soils on which conifers tend to grow may facilitate the higher proportion of apatite-Ca. Moreover, ectomycorrhizal fungi associated with the roots of conifers may be more effective at weathering apatite. In contrast, maple—the only major endomycorrhizal tree in the forest²⁵—is least able to access the Ca-apatite and is also the most sensitive species to Ca depletion in forest soils of the northeastern USA^{6,26}.

An alternative explanation for the higher Ca/Sr ratios in foliage compared to those in the soil exchange complex would require that trees preferentially take up Ca over Sr by a factor of 2.3 for maple, 2.9 for birch, 3.7 for beech and 8.5 for spruce/fir. We consider this unlikely for several reasons, but acknowledge that additional research on plant uptake of Ca and Sr is needed. First, to maintain mass balance, preferential uptake of Ca over Sr would leave the soil exchange complex depleted in Ca relative to Sr compared to silicate minerals, and this is not observed (Fig. 3). Second, plants grown in nutrient solutions display only minimal preferential uptake of Ca over Sr (by a factor of 1 to 1.4)¹⁸. A study of trees growing on

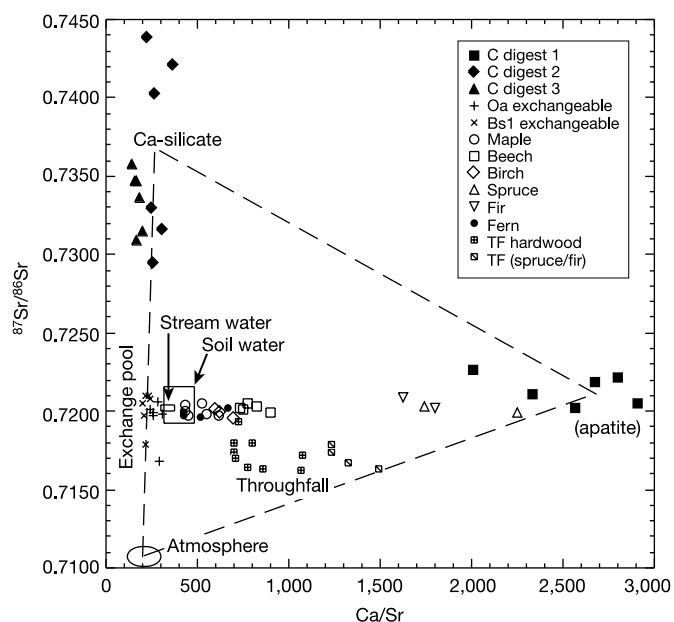


Figure 3 The Ca/Sr and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of all samples analysed. The small rectangular box represents the total variation in weekly streamwater samples collected during 1998, and the larger rectangular box represents the total variation in 28 soil water samples collected during each season in 1998. Dashed lines enclose the range of possible compositions for mixtures of Ca-silicate, apatite and atmospheric deposition. TF, throughfall.

carbonate bedrock suggested that spruce may assimilate Sr preferentially to Ca, whereas beech and maple do not display preferential uptake²⁷. Other studies of the uptake of Ca and Sr by trees growing on soils derived from silicate minerals show higher Ca/Sr of foliage compared to the exchange complex by a factor of 4.7 for beech²⁷ and 7.9–8.4 for spruce^{16,27}. We suggest that these values may reflect the presence or absence of apatite in the mineral soils, combined with rooting depth and mycorrhizal associations, rather than species-specific differences in plant uptake. Nevertheless, we consider the proportions of apatite-Ca in foliage given above to be maximum values, as some portion of the enhanced Ca/Sr may be due to preferential uptake of Ca over Sr.

Despite the importance of apatite-derived Ca to foliage, the Oa exchange complex contains only ~20% Ca from apatite. To shift the Ca/Sr ratio of the Bs1 exchange complex to the value of the Oa exchange complex, as much as 40% of the Ca would need to be supplied from maple litter or as little as 22% from spruce/fir litter. Most of the Ca in the Oa exchange complex (60–78%) is derived from a combination of silicate weathering and atmospheric deposition, in a 55:45 proportion. Ferns have Ca/Sr ratios consistent with maple foliage, suggesting that maple trees may access Ca from the forest floor without it entering an exchangeable form in the Oa horizon, as has been documented for P (ref. 28). In contrast, spruce and fir trees appear to extract Ca from apatite deep in the soil in addition to utilizing the forest floor pool.

We estimate that Ca in stream water is derived as follows: ~30% from the atmosphere, ~35% from silicate minerals, and ~35% from apatite. Previous research at HBEF^{1–3,7,29} suggested, on the basis of Ca/Na ratios, that silicate weathering is insufficient to account for Ca losses in stream water, and that Ca is being depleted from exchangeable and organically bound pools in forest soils of the northeastern USA. Our data suggest that apatite weathering, possibly by the direct action of root mycorrhizal systems, may account for a significant portion of this Ca, which has no accompanying Na release. However, this apatite pool itself has largely been depleted from upper soil horizons. Based on the average mass of the Bs2 horizon per hectare (ref. 30), and the amount of Ca released by the digest 1 of each soil horizon, we estimate the apatite Bs2 Ca reservoir to be 14 kmol ha⁻¹. This pool is larger than the entire exchangeable Ca pool of the soil (5 kmol ha⁻¹; ref. 2), and the forest floor Ca pool (8 kmol ha⁻¹; ref. 2), but is less than the pool of Ca in above-ground biomass (15 kmol ha⁻¹; ref. 2). Therefore, depletion of Ca from base-poor soils remains a considerable concern, but assessments and models of Ca depletion need to account for this important Ca source. □

Methods

Soil pits were excavated in 1997 at 47 sites in W-1 and sampled by horizon (Oa, E, Bh, Bs1, Bs2, C). Composite samples of the <2-mm size fraction were made for each horizon for six zones within W-1. Optical microscopy and electron microprobe analysis of soil parent material was used to determine the main Ca-bearing minerals and their compositions. Soils were subjected to a four-step sequential digestion on 0.5-g samples. Solutions were removed by centrifugation after each treatment, and are referred to in the main text as follows: 'exchangeable', 5 ml of 1 M NH₄Cl (pH = 7) for 20 h; 'digest 1', 5 ml of 1 M HNO₃ at room temperature for 20 h; 'digest 2', 5 ml of 1 M HNO₃ at 200 °C for 30 min (by microwave at ~20 MPa); 'digest 3', 10 ml of concentrated HF and HNO₃ at 210 °C for 30 min (by microwave at ~20 MPa). Composite samples of understory foliage (UF) and overstory foliage (OF) were collected in 1999 from trees encountered along transects through each of four elevation zones. Foliage sampling included red spruce (*Picea rubens*, 2 UF), balsam fir (*Abies balsamea*, 2 UF), American beech (*Fagus grandifolia*, 3 UF and 2 OF), yellow birch (*Betula alleghaniensis*, 3 OF), sugar maple (*Acer saccharum*, 3 UF and 3 OF) and wood fern (*Dryopteris spinulosa*, 4 UF). Throughfall was collected during a single storm event on 29 July 1999 at 13 locations in the watershed; soil water was collected using zero tension lysimeters during April, June, October and December of 1998 from the Oa, Bh and Bs1 horizons at three locations in the watershed; and weekly stream water samples from the base of W-1 were analysed for the 1998 calendar year. Foliage samples (0.5 g) were ground, homogenized and digested in concentrated HNO₃ at 200 °C for 30 min (by microwave at ~25 MPa). Leaches, digests and waters were analysed for Ca, K, Mg, Na, Si, P and Sr (±2%) and ⁸⁷Sr/⁸⁶Sr (±0.002%) following established methods^{12,14}.

Received 15 January; accepted 23 April 2002; doi:10.1038/nature00793.

- Likens, G. E., Driscoll, C. T. & Buso, D. C. Long-term effects of acid rain: Response and recovery of a forest ecosystem. *Science* **272**, 244–246 (1996).
- Likens, G. E. *et al.* The biogeochemistry of calcium at Hubbard Brook. *Biogeochemistry* **41**, 89–173 (1998).
- Bailey, S. W., Buso, D. C. & Likens, G. E. Implications of sodium mass balance for interpreting the calcium cycle of a northern hardwood ecosystem. *Ecology* (submitted).
- Huntington, T. G. The potential for calcium depletion in forest ecosystems of southeastern United States: Review and analysis. *Glob. Biogeochem. Cycles* **14**, 623–638 (2000).
- Federer, C. A. *et al.* Long-term depletion of calcium and other nutrients in eastern US forests. *Environ. Mgmt* **13**, 593–601 (1989).
- Lawrence, G. B. *et al.* Soil calcium status and the response of stream chemistry to changing acidic deposition rates in the Catskill Mountains of New York. *Ecol. Applic.* **9**, 1059–1072 (1999).
- Johnson, C. E., Driscoll, C. T., Siccama, T. G. & Likens, G. E. Element fluxes and landscape position in a northern hardwood forest watershed ecosystem. *Ecosystems* **3**, 159–184 (2000).
- Likens, G. E. & Borman, F. H. *Biogeochemistry of a Forested Ecosystem*, 2nd edn (Springer, New York, 1995).
- Jones, E. V. *et al.* The dissolution of apatite in the presence of aqueous metal cations at pH 2–7. *Chem. Geol.* **15**, 215–233 (1998).
- Brantley, S. L. & Chen, Y. in *Chemical Weathering Rates of Silicate Minerals* (eds White, A. F. & Brantley, S. L.) 119–172 (Mineralogical Society of America, Washington DC, 1995).
- Blum, A. E. & Stillings, L. L. in *Chemical Weathering Rates of Silicate Minerals* (eds White, A. F. & Brantley, S. L.) 291–352 (Mineralogical Society of America, Washington DC, 1995).
- Hogan, J. E., Blum, J. D., Siegel, D. I. & Glaser, P. H. ⁸⁷Sr/⁸⁶Sr as a tracer of groundwater discharge and precipitation recharge in the Glacial Lake Agassiz Peatlands, Northern Minnesota, USA. *Wat. Resour. Res.* **36**, 3701–3710 (2000).
- Capo, R. C., Stewart, B. W. & Chadwick, O. A. Strontium isotopes as tracers of ecosystem processes: theory and methods. *Geoderma* **82**, 197–225 (1998).
- Blum, J. D., Taliaferro, H., Weisse, M. T. & Holmes, R. T. Changes in Sr/Ca, Ba/Ca and ⁸⁷Sr/⁸⁶Sr ratios between trophic levels in two forest ecosystems in the northeastern USA. *Biogeochemistry* **49**, 87–101 (2000).
- Miller, E. K., Blum, J. D. & Friedland, A. J. Determination of soil exchangeable-cation loss and weathering rates using Sr isotopes. *Nature* **362**, 438–441 (1993).
- Bailey, S. W., Hornbeck, J. W., Driscoll, C. T. & Gaudette, H. E. Calcium inputs and transport in a base-poor forest ecosystem as interpreted by Sr isotopes. *Wat. Resour. Res.* **32**, 707–719 (1996).
- Vitousek, P. M., Kennedy, M. J., Derry, L. A. & Chadwick, O. A. Weathering versus atmospheric sources of strontium in ecosystems on young volcanic soils. *Oecologia* **121**, 255–259 (1999).
- Rumia, L. T. Strontium and calcium distribution in plants: effect on palaeodietary studies. *J. Archaeol. Sci.* **14**, 599–608 (1987).
- Aberg, G., Jacks, G. & Hamilton, J. Weathering rates and ⁸⁷Sr/⁸⁶Sr ratios: an isotopic approach. *J. Hydrol.* **109**, 65–78 (1989).
- Kistler, R. W., Chappell, B. W., Peck, D. L. & Bateman, P. C. Isotopic variation in the Tuolumne Intrusive Suite, central Sierra Nevada, California. *Contrib. Mineral. Petrol.* **94**, 205–220 (1986).
- Creaser, R. A. & Gray, C. M. Preserved initial Sr-87/Sr-86 in apatite from altered felsic igneous rocks—a case study from the middle Proterozoic of south-Australia. *Geochim. Cosmochim. Acta.* **56**, 2789–2795 (1992).
- Van Breeman, N. *et al.* Mycorrhizal weathering: a true case of mineral plant nutrition? *Biogeochemistry* **49**, 53–67 (2000).
- Wallander, H., Johansson, L. & Pallon, J. PIXE analysis to estimate the elemental composition of ectomycorrhizal rhizomorphs grown in contact with different minerals in forest soil. *FEMS Microbiol. Ecol.* **1320**, 1–10 (2002).
- Wallander, H. Uptake of P from apatite by *Pinus sylvestris* seedlings colonized by different ectomycorrhizal fungi. *Plant Soil* **218**, 249–256 (2000).
- Yawney, W. J. & Schultz, R. C. Anatomy of a vesicular-arbuscular endomycorrhizal symbiosis between sugar maple (*Acer saccharum* Marsh) and *Glomus etunicatum* Becker and Gerdemann. *New Phytol.* **114**, 47–57 (1990).
- Long, R. P., Horsley, S. B. & Lilja, P. R. Impact of forest liming on growth and crown vigor of sugar maple and associated hardwoods. *Can. J. Forest. Res.* **27**, 1560–1573 (1997).
- Pozwa, A., Dambrine, E., Pollier, B. & Attia, O. A comparison between Ca and Sr cycling in forest ecosystems. *Plant Soil* **225**, 299–310 (2000).
- Yanai, R. D. Phosphorus budget of a 70-year-old northern hardwood forest. *Biogeochemistry* **17**, 1–22 (1992).
- Driscoll, C. T. *et al.* Acidic deposition in the northeastern United States: sources and inputs, ecosystem effects, and management strategies. *Bioscience* **51**, 180–198 (2001).
- Johnson, C. E. & Petras, R. J. Distribution of zinc and lead fractions within a forest Spodosol. *Soil Sci. Soc. Am. J.* **62**, 782–789 (1998).

Acknowledgements

We thank S. Bailey, D. Buso, S. Hamburg, J. Hogan and R. Yanai for discussions, and H. Wallander and L. Derry for comments and suggestions. This research was supported by the NSF, and carried out at the Hubbard Brook experimental forest, which is operated and maintained by the Northeastern Forest Experiment Station, US Department of Agriculture.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.D.B. (e-mail: jdblum@umich.edu).

Density-dependent mortality and the latitudinal gradient in species diversity

Janneke Hille Ris Lambers^{*†}, James S. Clark^{*‡} & Brian Beckage^{*†}

^{*} Biology Department, Duke University, Durham, North Carolina 27708, USA

[‡] Nicholas School of the Environment and Earth Sciences, Durham, North Carolina 27708, USA

[†] Present addresses: Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul, Minnesota 55108, USA (J.H.R.L.); Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996, USA (B.B.)

Ecologists have long postulated that density-dependent mortality maintains high tree diversity in the tropics^{1–6}. If species experience greater mortality when abundant, then more rare species can persist^{1,2,7–9}. Agents of density-dependent mortality (such as host-specific predators, and pathogens) may be more prevalent or have stronger effects in tropical forests, because they are not limited by climatic factors^{1–5}. If so, decreasing density-dependent mortality with increasing latitude could partially explain the observed latitudinal gradient in tree diversity^{4–6}. This hypothesis has never been tested with latitudinal data. Here we show that several temperate tree species experience density-dependent mortality between seed dispersal and seedling establishment. The proportion of species affected is equivalent to that in tropical forests^{6,10–16}, failing to support the hypothesis that this mechanism is more prevalent at tropical latitudes. We further show that density-dependent mortality is misinterpreted in previous studies. Our results and evidence from other studies suggest that density-dependent mortality is important in many forests. Thus, unless the strength of density-dependent mortality varies with latitude, this mechanism is not likely to explain the high diversity of tropical forests.

We used a demographic study of temperate forest species to determine whether survival probabilities at early life-history transitions decrease when seeds or seedlings are at high densities or close to adults of the same species (conspecific). We quantified seed dispersal, seed bank densities, seedling emergence and seedling survival at 100 microsites distributed across five temperate forest stands over 3 years. We then determined whether transition probabilities across early life-history stages were lower at high conspecific seed and seedling densities or close to conspecific trees, a phenomenon said to promote species diversity, called the 'Janzen–Connell' effect^{1,2}. We tested transitions from dispersed seed to seed bank (incorporation), seed bank to germinated seedling (germination), germinated seedling to established seedling (establishment), and first-year seedling to second-year seedling (survival). We also determined whether effects of conspecific density (seeds, seedlings) or distance (proximity to conspecific trees) on mortality are greater than are effects of other species, a requirement if density-dependent mortality is to promote species diversity^{1,9}.

Six of the seven temperate forest taxa we tested (*Acer pennsylvanicum*, *Acer rubrum*, *Betula* spp., *Fraxinus americana*, *Liriodendron tulipifera* and *Quercus rubra*) experience density-dependent mortality at one or more life-history transitions, with greater effects of conspecific seed and seedling density or conspecific adult proximity on mortality than of heterospecific (different species) (Table 1). Survival of seeds in the soil (three of three species with long-term seed banks) and seedling emergence (five of seven species) was reduced at high conspecific seed densities, close to conspecific trees, or both. For example, the proportion of *Acer rubrum* seeds germinating in 1 × 1 m plots is depressed at high seed rain densities (Table 1, Fig. 1). Because the negative effects of conspecifics are much greater than those of heterospecifics (Table 1), they are likely to result from host-specific seed predators or pathogens^{17–21} rather than resource competition with other seedlings or adults. Regardless of what causes seedling mortality, our analyses identified the kind of density-dependent mortality that promotes species diversity by disproportionately affecting species when they are abundant^{10,15}.

The high numbers of tree species experiencing density-depen-

Table 1 Parameter estimates and probabilities from best-fitting models

Species Transition	Parameters (from equation 2)					P-values			
	Intercept v_j	Conspecific effects		Heterospecific effects		Density-dependent mortality		Conspecific > heterospecific effects	
		Juveniles d_j	Adults a_j	Juveniles d	Adults a	($d_j = 0$)	($a_j = 0$)	($d_j = d$)	($a_j = a$)
<i>Acer pennsylvanicum</i>									
Germination	1.54×10^{-1}	8.27×10^{-2}	–	1.26×10^{-5}	–	<0.001	NS	<0.001	NS
Establishment	2.24×10^{-1}	–	-5.91×10^0	–	-8.71×10^{-1}	NS	NS	NS	NS
Survival	7.33×10^{-1}	6.00×10^{-2}	-5.17×10^0	6.00×10^{-2}	-3.04×10^{-1}	0.028	0.022	NS	0.027
<i>Acer rubrum</i>									
Incorporation	1.29×10^{-1}	4.57×10^{-2}	–	3.71×10^{-4}	–	<0.001	NS	<0.001	NS
Germination	3.07×10^{-1}	5.15×10^{-2}	5.36×10^{-1}	2.77×10^{-5}	5.36×10^{-1}	<0.001	0.020	<0.001	NS
Establishment	1.91×10^{-1}	-1.38×10^{-2}	–	-1.38×10^{-1}	–	<0.001	NS	NS	NS
Survival	4.49×10^{-1}	-1.10×10^{-2}	–	1.18×10^{-2}	–	<0.001	NS	<0.001	NS
<i>Betula</i> spp.									
Incorporation	2.97×10^{-2}	2.92×10^{-4}	1.14×10^1	-2.90×10^{-3}	-8.26×10^{-1}	<0.001	<0.001	<0.001	<0.001
Germination	1.79×10^{-3}	3.38×10^{-3}	–	-1.38×10^{-3}	–	<0.001	NS	0.002	NS
Establishment	9.89×10^{-1}	–	8.76×10^0	–	8.76×10^0	NS	0.007	NS	NS
Survival	1.03×10^{-4}	4.88×10^{-2}	-4.06×10^1	4.88×10^{-2}	-4.06×10^1	0.004	0.029	NS	NS
<i>Fraxinus americana</i>									
Germination	1.760×10^{-1}	2.61×10^{-1}	–	5.23×10^{-4}	–	0.008	NS	0.009	NS
Establishment	3.79×10^{-1}	–	–	–	–	NS	NS	NS	NS
<i>Liriodendron tulipifera</i>									
Incorporation	9.99×10^{-2}	1.26×10^{-2}	-5.97×10^0	5.77×10^{-4}	-1.01×10^{-1}	<0.001	0.002	<0.001	0.003
Germination	2.66×10^{-3}	6.89×10^{-3}	-6.49×10^0	6.89×10^{-3}	4.51×10^0	0.039	NS	NS	0.008
Establishment	3.99×10^{-1}	2.53×10^{-2}	–	2.53×10^{-2}	–	0.007	NS	NS	NS
<i>Quercus rubra</i>									
Germination	1.89×10^{-2}	9.16×10^{-3}	3.84×10^0	9.16×10^{-3}	-5.54×10^0	<0.001	NS	NS	0.008
Establishment	7.64×10^{-1}	–	–	–	–	NS	NS	NS	NS
Survival	4.29×10^{-1}	–	–	–	–	NS	NS	NS	NS
<i>Vitis</i> spp.									
Germination	4.70×10^{-4}	–	–	–	–	NS	NS	NS	NS

A dash indicates that the parameter not significantly different from 0, as determined from likelihood ratio tests. NS, $P > 0.05$, not significant.

dent mortality in the tropics is sometimes hypothesized to explain the latitudinal gradient in species diversity^{4–6}. Our results showing high levels of density-dependent mortality in temperate forests (and studies in Texas²² and South Carolina²³) led us to examine the relationship between latitude and the proportion of tree species experiencing density-dependent mortality from ten studies in eight forest communities (two in Australia^{10,14}, two in Indonesia^{12,15,16}, one in Panama^{6,11,13,16}, and three in southeastern USA—Texas²², South Carolina²³ and our site in North Carolina). We found no

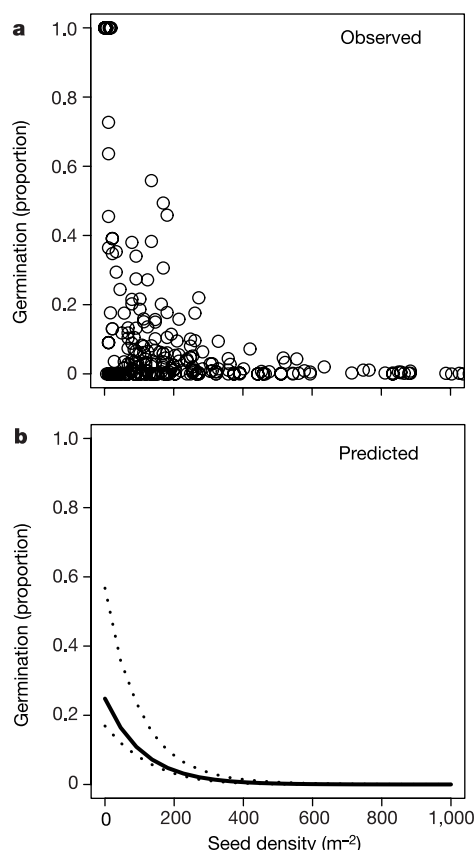


Figure 1 The observed (a) and predicted (b) relationship between conspecific seed density and seed germination for *Acer rubrum*. **a**, Each circle represents a paired seed trap and seedling quadrat in 1 of 3 years. **b**, The best-fit relationship between germination and density of conspecific seeds (see Methods); dashed lines are 95% bootstrapped confidence intervals. Likelihood ratio tests indicate that germination decreases at high conspecific densities.

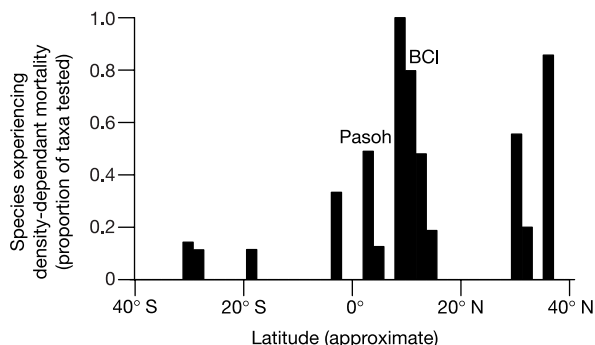


Figure 2 The proportion of tree taxa experiencing density-dependent mortality at different latitudes. Each bar corresponds to an analysis testing a community of forest trees for the presence of density-dependent mortality (our analyses and previous studies^{6,10–16,22,23}). The proportion of the forest community experiencing density-dependent mortality does not increase at tropical latitudes.

evidence that the proportion of species experiencing density-dependent mortality is correlated with latitude (Spearman's rank correlation = 0.028, $P > 0.9$). Thus, although the magnitude of density-dependent mortality (that is, strength of effects) could be higher in tropical forests (we have no evidence that it is), the prevalence of density-dependent mortality (that is, number of species affected) is not associated with latitude.

The role of density dependence in promoting forest tree diversity has been extensively debated^{8,11,12,24–29}. Empirical studies have not resolved this question, because the number of species interpreted as experiencing density-dependent mortality varies among studies, even within the same forest (for example, studies at Pasoh^{12,13} and Barro Colorado Island^{6,11,13,16}; Fig. 2). We re-analysed our data using assumptions employed by studies reporting variable levels of density-dependent mortality in forests^{6,11–14,16} and found that the different results may reflect methods of analysis rather than underlying differences in the prevalence of density-dependent mortality (see Methods and Fig. 3).

Using methods employed in previous studies, we might have concluded that all species at all life-history stages experience density-dependent mortality (Fig. 3e) or that no species experience density-dependent mortality at any life-history stage (Fig. 3d). Density-dependent mortality at early life-history stages may be underestimated if only effects of adult trees^{11,12} are quantified, that is, if negative effects of juvenile conspecifics are ignored (compare Fig. 3a, b). In contrast, failing to distinguish between con- and heterospecific effects^{6,12,29} can lead to overestimation of density-dependent mortality (compare Fig. 3a, c), because effects must be species-specific to foster coexistence of species^{1,9}. Owing to dispersal, initial seed densities are highest around parent trees, and often remain so in spite of high mortality at high densities of seeds or seedlings or close to parent trees. Analyses that ignore dispersal by testing whether densities of recruits increase with distance from conspecific trees^{12,14} can therefore underestimate density-dependent mortality (compare Fig. 3a, d). Statistics used to indicate density-dependent mortality in one study⁶ overestimate the prevalence of this mechanism relative to other methods (compare Fig. 3a–d with e). Finally, more species experience density-dependent mortality at the earliest life-history stages², regardless of methodology (Fig. 3). Studies performed on later life-history transitions^{11–13,16} thus underestimate the number of taxa affected by density-dependent mortality. Overall, the prevalence of density-dependent mortality has probably been underestimated in forests, because density-dependent mortality is most prevalent at seed and seedling stages (Fig. 3), and most studies reporting low density-dependent mortality are performed at later life-history stages^{11–13,16}.

High density-dependent mortality in temperate forests (Table 1) and the lack of a latitudinal gradient in density-dependent effects (Fig. 2) do not support the view that density-dependent mortality is responsible for the latitudinal gradient in species diversity. We find that density-dependent mortality is as common in temperate forests as in tropical forests (as do other studies^{21–23}). Confusion has resulted from analyses that ignore dispersal or are applied at later life-history stages, both of which underestimate density-dependent mortality (Fig. 3). The ubiquity of density-dependent mortality may mean this mechanism contributes to tree diversity in forests worldwide, if effects at early life-history stages promote diversity at later life-history stages (with one exception⁶, rarely tested). Absence of evidence for latitudinal trends in density-dependent mortality implies that other mechanisms are responsible for high diversity in the tropics. □

Methods

Study site and experimental design

Studies were performed in secondary temperate deciduous forests at the Coweeta Hydrologic Laboratory (western North Carolina). We identified and mapped all trees more than 2 m tall in five 80 × 80-m permanent vegetation plots³⁰. Twenty seed traps were

located within each plot. Seed traps were mesh-lined baskets (0.1764 m^2) suspended 1 m above the ground and covered with wire mesh to deter seed predation³⁰. Traps were emptied four times annually, and seeds identified and counted³⁰. Soil cores (0.0062 m^2 , 0.05 m deep) adjacent to each trap were removed in autumn 1996, 1997 and 1998. In Duke University's greenhouses, soil was spread over sterile potting soil, and seedlings (each representing a viable seed) were counted and identified as they emerged. In the field, seedlings were identified and tagged in during summer 1997, 1998 and 1999 in quadrats (1.0 m^2) adjacent to each seed trap.

For seven species having sufficient densities to perform analyses, we tested whether four life-history transitions (incorporation, germination, establishment and survival) are affected by conspecific juvenile densities or adult proximity. Analyses were performed on data from 3 years, except for seed incorporation into the soil, where data were averaged across years (because seeds survive more than 1 year in the soil). Results were consistent when data were analysed as averages across years, or separately by plot and by year.

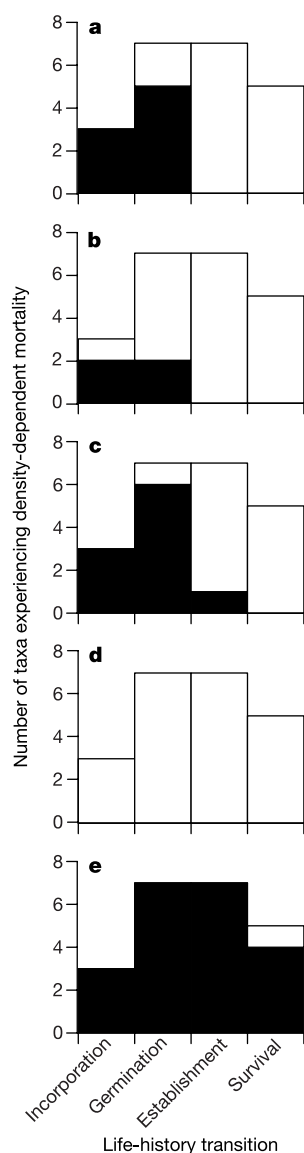


Figure 3 Five analyses used to test for density-dependent mortality are applied to seed bank incorporation, germination, establishment and over-winter survival. **a**, Our analyses (equations 1 and 2). **b**, Tests that ignore density-dependent mortality due to high juvenile densities. **c**, Tests that do not differentiate conspecific density-dependent mortality from heterospecific density-dependent mortality. **d**, Tests that ignore effects of dispersal on seed and seedling distributions (equation (3)). **e**, Regression of logged data to test whether survival decreases at high densities (equations (4), (5) and (6)). Both analyses employed and life-history transitions tested strongly influence conclusions about the prevalence of density-dependent mortality.

Estimation and testing

For transitions from life-history stage $k - 1$ to k for species j , we used the binomial likelihood to estimate transition probabilities (P_{ijk}) and to test for density-dependent mortality in T quadrats:

$$L_{jk} = \prod_{i=1}^T \text{Bin}(n_{ijk} | n_{ij(k-1)}, P_{ijk}(v_{jk}, d_{jk}, a_{jk}, a_k)) \quad (1)$$

The probability of survival of species j at quadrat i ,

$$P_{ijk} = v_{jk} \exp \left[-d_{jk} n_{ij(k-1)} - d_k \sum_{m \neq j} n_{im(k-1)} - a_{jk} \sum_{ij} b_j / x_{ij} - a_k \sum_{m \neq j} b_m / x_{im} \right] \quad (2)$$

depends on conspecific juvenile densities at quadrat i ($n_{ij(k-1)}$) and juvenile densities of all other species ($n_{im(k-1)}$). Density dependence means that fitted parameters d_{jk} and d_k are greater than zero. If transition probabilities are decreased close to surrounding adult trees (basal area b_j at distance x_{ij} metres), fitted parameters a_{jk} and a_k are greater than zero. A more traditional analysis of distance-dependent mortality² did not change our results, and led to poorer model fits (as determined by Akaike's information criterion). The transition probability tends to v_{jk} at low juvenile densities, far from adult trees, or when there are no effects of juveniles or adults.

We used likelihood ratio tests to determine whether conspecific juveniles have stronger effects on transition probabilities than other species ($d_{jk} > d_k$), or whether mortality is more strongly affected by conspecific trees than by heterospecific trees ($a_{jk} > a_k$). In the few cases where seed bank or seedling densities within quadrats were greater than incoming seed densities in neighbouring seed traps, we assumed seed densities were equivalent to seed bank or seedling densities for these quadrats. We repeated analyses excluding those sampling sites, and found consistent results.

Comparisons with other studies

To determine whether methodology accounts for differences among studies in the proportion of taxa affected by density-dependent mortality, we compared the number of taxa experiencing density-dependent mortality using likelihood ratio tests and equations 1 and 2 (Fig. 3a) with those obtained using four approaches in previous studies (Fig. 3b–e).

The first analysis assumes that only adult trees cause density-dependent mortality. We re-analysed our data in a comparable fashion by assuming no effects of juveniles on P_{ijk} , by setting d_{jk} and d_k in equation 2 to zero (Fig. 3b). The second analysis assumes that there are no heterospecific effects (either adult or juvenile) on early life-history transitions. For this analysis, we set both d_{jk} and a_{jk} in equation 2 to zero (Fig. 3c). The third method assumes that mortality close to conspecific trees depresses recruitment, and that seedling densities increase with distance from trees. We examined this assumption by testing whether seed, seed bank or seedling densities (n_{ijk}) were positively correlated with the distance of each quadrat (x_{ij}) to its nearest conspecific tree (fitted parameter $b_{jk} > 0$):

$$n_{ijk} = a_{jk} + b_{jk} x_{ij} \quad (3)$$

using a negative binomial likelihood (Fig. 3d). This test is analogous to that made in several studies^{12,14} of density-dependent mortality, although we used an error distribution more appropriate for count data.

The fourth method tests whether juvenile densities (n_{ijk}) saturate at high inputs ($n_{ij(k-1)}$):

$$n_{ijk} = a \times (n_{ij(k-1)})^b \quad (4)$$

Fitted parameters are b and c . We fitted this relationship in a manner similar to that in a recent study⁶—with a linear regression using values logged after adding one:

$$\ln(n_{ijk} + 1) = \ln b + c \ln(n_{ij(k-1)} + 1) + \varepsilon_{ij} \quad (5)$$

with normally (N) distributed error:

$$\varepsilon_{ijk} \sim N(0, \sigma^2) \quad (6)$$

With this analysis, density-dependent mortality is indicated by a slope parameter, c , being less than one (Fig. 3e). The addition of a constant to juvenile densities n_{ijk} (in this case, one) before log transformation underestimates the slope parameter (c), potentially overestimating density-dependent mortality.

Received 3 October 2001; accepted 8 March 2002; doi:10.1038/nature00809.

- Janzen, D. H. Herbivores and the number of tree species in tropical forests. *Am. Nat.* **104**, 501–528 (1970).
- Connell, J. H. In *Dynamics of Populations* (eds den Boer, P. J. & Gradwell, G. R.) 298–312 (Center for Agricultural Publication and Documentation, Wageningen, The Netherlands, 1971).
- Coley, P. D. & Barone, J. A. Herbivory and plant defenses in tropical forests. *Annu. Rev. Ecol. Syst.* **27**, 305–335 (1996).
- Givnish, T. J. On the causes of gradients in tropical tree diversity. *J. Ecol.* **87**, 193–210 (1999).
- Leigh, E. G. *Tropical Forest Ecology* 190–195 (Oxford University Press, New York, 1999).
- Harms, K. E., Wright, S. J., Calderon, O., Hernandez, A. & Herre, E. A. Pervasive density-dependent recruitment enhances seedling diversity in a tropical forest. *Nature* **404**, 493–495 (2000).
- Howe, H. F. & Smallwood, J. Ecology of seed dispersal. *Annu. Rev. Ecol. Syst.* **13**, 201–228 (1982).
- Becker, P., Lee, L. W., Rotham, E. D. & Hamilton, W. D. Seed predation and the coexistence of tree species: Hubbell's models revisited. *Oikos* **44**, 382–390 (1985).
- Pacala, S. W. & Crawley, M. J. Herbivores and plant diversity. *Am. Nat.* **140**, 243–260 (1992).
- Connell, J. H., Tracey, J. G. & Webb, L. J. Compensatory recruitment, growth, and mortality as factors maintaining rain forest tree diversity. *Ecol. Monogr.* **54**, 141–164 (1984).
- Condit, R., Hubbell, S. P. & Foster, R. B. Recruitment near conspecific adults and the maintenance of tree and shrub diversity in a neotropical forest. *Am. Nat.* **140**, 261–286 (1992).
- Okuda, T., Kachi, N., Yap, S. K. & Manokaran, N. Tree distribution pattern and fate of juveniles in a lowland tropical rain forest—implications for regeneration and maintenance of species diversity. *Plant*

- Ecol.* **131**, 155–171 (1997).
13. Wills, C. & Condit, R. Similar non-random processes maintain diversity in two tropical rainforests. *Proc. R. Soc. Lond. B* **266**, 1445–1452 (1999).
 14. Penfold, G. C. & Lamb, D. Species co-existence in an Australian subtropical rain forest: evidence for compensatory mortality. *J. Ecol.* **87**, 316–329 (1999).
 15. Webb, C. O. & Peart, D. R. Seedling density dependence promotes coexistence of Bornean rain forest trees. *Ecology* **80**, 2006–2017 (1999).
 16. Wills, C., Condit, R., Foster, R. B. & Hubbell, S. P. Strong density- and diversity-related effects help to maintain tree species diversity in a neotropical forest. *Proc. Natl Acad. Sci. USA* **94**, 1252–1257 (1997).
 17. Augsburger, C. K. Seedling survival of tropical tree species: interactions of dispersal distance, light gaps, and pathogens. *Ecology* **65**, 1705–1712 (1984).
 18. Gilbert, G. S., Hubbell, S. P. & Foster, R. B. Density and distance to adult effects of a canker disease of trees in moist tropical forest. *Oecologia* **98**, 100–108 (1994).
 19. Cintra, R. A test of the Janzen–Connell model with two common tree species in Amazonian forest. *J. Trop. Ecol.* **8**, 529–536 (1997).
 20. Carson, W. P. & Root, R. R. Herbivory and plant species coexistence: community regulation by an outbreaking phytophagous insect. *Ecol. Monogr.* **70**, 73–99 (2000).
 21. Packer, A. & Clay, K. Soil pathogens and spatial patterns of seedling mortality in a temperate tree. *Nature* **404**, 278–281 (2000).
 22. Streng, D. R., Glitzenstein, J. S. & Harcombe, P. A. Woody seedling dynamics in an east Texas floodplain forest. *Ecol. Monogr.* **59**, 177–204 (1989).
 23. Jones, R. H., Sharitz, R. R., Dixon, P. M., Segal, D. S. & Schneider, R. L. Woody plant regeneration in four floodplain forests. *Ecol. Monogr.* **64**, 345–367 (1994).
 24. Connell, J. H. Diversity in tropical rain forests and coral reefs. *Science* **199**, 1302–1310 (1978).
 25. Hubbell, S. P. Seed predation and the coexistence of tree species in tropical forests. *Oikos* **35**, 214–229 (1980).
 26. McCanny, S. J. Alternatives in parent–offspring relationships in plants. *Oikos* **45**, 148–149 (1985).
 27. Hubbell, S. P., Condit, R. & Foster, R. B. Presence and absence of density dependence in a neotropical tree community. *Phil. Trans. R. Soc. Lond. B* **330**, 269–281 (1990).
 28. Schupp, E. W. The Janzen–Connell model for tropical tree diversity: population implications and the importance of spatial scale. *Am. Nat.* **140**, 526–530 (1992).
 29. Clark, D. A. & Clark, D. B. Spacing dynamics of a tropical rain forest tree: evaluation of the Janzen–Connell model. *Am. Nat.* **124**, 769–788 (1984).
 30. Clark, J. S., Macklin, E. & Wood, L. Stages and spatial scales of recruitment limitation in southern Appalachian forests. *Ecol. Monogr.* **68**, 213–235 (1998).

Acknowledgements

Funding was provided by Sigma Xi and National Science Foundation grants. We thank M. Dietze, K. Harms, I. Ibanez, S. Ladeau, J. Lynch, J. McLachlan, J. Mohan, A. Pringle and M. Rocca for discussion of data and comments on drafts of the manuscript. Support in the field was provided by R. Hille Ris Lambers and H. Passmore.

Competing interests statement

The authors declare that they have no competing financial interests

Correspondence and requests for materials should be addressed to J.H.R.L. (e-mail: hille015@umn.edu).

Geographic structure and dynamics of coevolutionary selection

John N. Thompson* & Bradley M. Cunningham†

* Department of Ecology and Evolutionary Biology, Earth and Marine Sciences Building, University of California, Santa Cruz, California 95064, USA

† School of Life Sciences, Washington State University, Pullman, Washington 99164, USA

Coevolution of species is one of the major processes organizing the Earth's biodiversity. Recent coevolutionary theory has indicated that the geographic structure of species has the potential to impose powerful and continuing effects on coevolutionary dynamics, if that structure creates selection mosaics and coevolutionary hotspots across landscapes^{1–7}. Here we confirm that current coevolutionary selection in interspecific interactions can be highly divergent across both narrow and broad geographic scales, thereby fuelling continuing coevolution of taxa. Study of a widespread plant–insect interaction across a broad range of habitats for several years showed that an insect functioning both as a pollinator and a floral parasite can be strongly

mutualistic in some habitats but commensal or antagonistic in neighbouring habitats. The results for one of the habitats span seven years, demonstrating that the local structure of coevolutionary selection can remain stable across multiple generations. Conservation of the evolutionary processes maintaining long-term biological diversity may require preservation of the conditions that allow a long-term shifting geographic mosaic of coevolutionary hotspots and coldspots.

Although some coevolving interactions show little spatial variation in outcome⁸, evidence for pronounced geographic differences in interspecific interactions has now been found in a diverse array of interactions, including those between parasites and hosts, predators and prey, competitors, and mutualists^{9–18}. The geographic mosaic theory of coevolution indicates that these geographic differences are an inherent part of the coevolutionary process, driven by variation between habitats in the trajectories of natural selection (selection mosaics), the occurrence of reciprocal selection in only some communities (coevolutionary hotspots), and a constantly changing

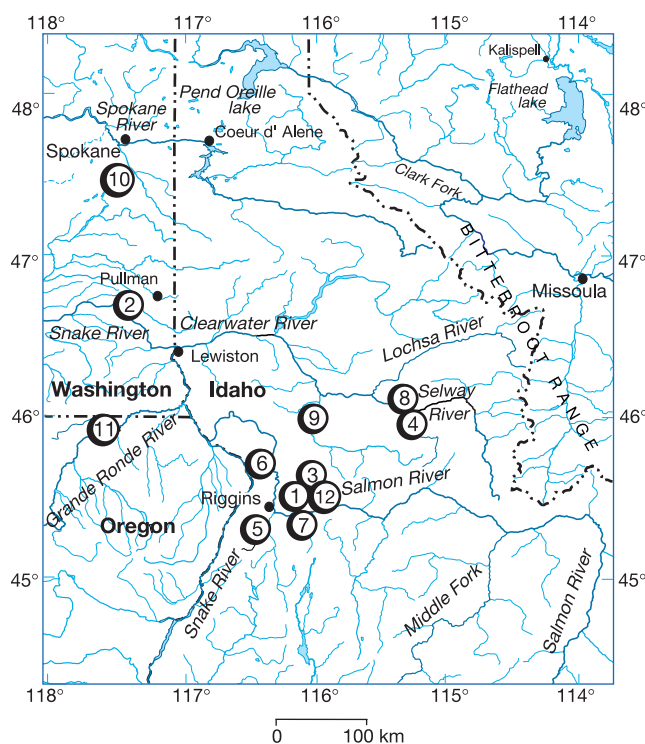


Figure 1 Distribution of study sites for analysis of coevolutionary hotspots in the interaction between the plant *Lithophragma parviflorum* and the moth *Greya politella*. Locations: 1, Berg = Berg Mountain, Salmon River, Idaho (ponderosa pine woodland); 2, Granite = Granite Point, Snake River, Washington (canyon bottom meadow steppe); 3, Keating = Keating Ridge, above Salmon River, Idaho (ridgetop grassland with ponderosa pine); 4, Meadow = Meadow Creek near confluence with Selway River, Idaho (streamside meadow amid grand fir and western red cedar); 5, Rapid = Rapid River above confluence with Little Salmon River, Idaho (streamside narrow canyon amid Douglas fir and Pacific yew); 6, Saddle = near Pittsburg Saddle above Snake River, Idaho (high-elevation open ridge top with scattered ponderosa pines); 7, Salmon = along Salmon River east of Riggins, Idaho (north-facing steppe); 8, Selway = lower Selway River, Idaho (openings in habitat dominated by western red cedar, Douglas fir and ponderosa pine); 9, South Fork = South Fork of Clearwater River, Idaho (open ponderosa pine savanna); 10, Turnbull = Turnbull National Wildlife Refuge, Washington (ponderosa pine woodland on glacial polygonal soil); 11, Wenaha = Wenaha River, Idaho (south-facing steppe with thick shrub cover); 12, Wind = Gospel Hump Wilderness along Salmon River, Idaho (dry, south-facing steppe).

genetic landscape shaped by gene flow and other evolutionary forces^{1,19}. Although coevolutionary hotspots and coldspots have now been demonstrated in a few interactions on the basis of the geographic distribution of coevolved traits^{11,12}, the distribution of current coevolutionary selection is unknown for any interaction. The absence of such data precludes the development of more ecologically grounded models of coevolutionary dynamics.

We evaluated current coevolutionary selection in the interaction between the herbaceous plant *Lithophragma parviflorum* (Saxifragaceae) and the floral-parasitic moth *Greya politella*^{20,21}. This is one of the most widespread plant–insect interactions in western North America and is pervasive in the Rocky Mountains, northern USA, a floristically diverse region that includes large expanses of relatively pristine environments shaped by rugged topography and Pleistocene events^{22,23}. Extensive field surveys between 1994 and 1999 within the major river drainages of western Idaho and adjacent areas of Washington and Oregon showed that all sampled *L. parviflorum* populations support *Greya* moth populations (Table 1). Moreover, the interaction is common within populations, with more than 25% of plants receiving eggs in most populations and years. At Turnbull National Wildlife Refuge in Washington State, where the local *L. parviflorum* population includes hundreds of thousands of individuals, 85–98% of plants received eggs during three years of study.

In all surveyed habitats, moths oviposited into developing flowers by inserting the ovipositor down the corolla and into the floral ovary. The moths are very close relatives of yucca moths²⁴ and, like those moths, pollinate their obligately outcrossing host flowers during oviposition. Unlike the yucca moths, however, pollination by *Greya* is passive rather than active, and occurs as pollen adhering to the abdomen touches the stigma during oviposition. Because the moths leave direct evidence of their visits (that is, eggs within flowers), we were able to assess the outcome of this pairwise interaction, even when the same plants were visited by co-pollinators, by comparing flowers with and without *Greya* eggs. Consequently, unlike most interactions, it was possible to evaluate this pairwise interaction within the full community context in which it occurs in natural populations.

In most surveyed habitats, the moths were restricted to feeding on *L. parviflorum*, both as larvae and as adults. Hence, the interaction is inherently beneficial to the moths in these communities. However, previous work^{20,21} and the preliminary observations during the surveys indicated that the interaction is at least sometimes less obligate for the plants, because co-pollinators visit the plants in some habitats, thereby potentially reducing *Lithophragma*'s reliance on *Greya* for pollination. Unlike the co-pollinators, the larvae eat a small subset of the hundreds of developing seeds within each capsule. Hence, the overall outcome of the interaction has the potential to be mutualistic, commensal or antagonistic, depending on the presence and abundance of effective co-pollinators in different habitats.

Table 2 Effect of moth oviposition on probability of floral development in *L. parviflorum*

Population	Effect of <i>Greya</i> oviposition on floral development		
	1997	1998	1999
Mutualistic effect of <i>Greya</i>			
Berg	1.1	3.0	1.8
Meadow	—	2.4	1.6
Saddle	0.8	20.5	9.2
Turnbull	1.0	1.0	2.4
Antagonistic effect of <i>Greya</i>			
Rapid	0.5	0.9	1.3
Salmon	—	0.4	—
South Fork	—	0.6	0.4
Wenaha	2.1	0.4	1.6
No effect of <i>Greya</i>			
Granite	0.5	1.1	1.0
Keating	1.1	1.1	1.2
Selway	—	0.9	—
Wind	—	0.7	1.0

Results are shown as the ratio of the percentage of developed capsules containing eggs to the percentage of aborted capsules containing eggs. Significant effects ($P < 0.05$) determined by χ^2 tests are shown in bold. Values significantly >1 indicate selective development of capsules containing eggs. Values significantly <1 indicate selective abortion of capsules containing eggs. The number of dissected capsules for each analysis averaged 678 (s.d. = 89.5, range = 386–898). Populations indicated with a dash were either not studied in that year or had too few aborted capsules for statistical analysis. Analyses were restricted to floral positions 2–4 on each plant.

We evaluated whether *L. parviflorum* relied upon *Greya* for pollination in 12 contrasting habitats ($N = 18,912$ flowers in total), ranging from dry open steppe to ponderosa pine (*Pinus ponderosa*) woodlands to steep basaltic or granitic slopes along moist river drainages (Fig. 1). In all except one of the habitats (Rapid River), the moths depend completely on *L. parviflorum* for adult and larval food and for mating sites. In each community we tested whether flowers visited by the moths were more likely to develop seeds than those not visited by the moths.

The interaction varied among habitats from mutualism to commensalism to antagonism. *Lithophragma parviflorum* relied on *G. politella* for pollination within four of the habitats in one or more years (Table 2). These four habitats provide evidence for current mutualistic hotspots: communities in which the moths depend completely on this one host plant and in which the plants depend upon the moths for pollination. Developed capsules at these sites were on average 1.8–20.5-fold more likely to have *Greya* eggs than were aborted capsules. The most extreme dependence of *Lithophragma* on *Greya* for pollination occurred at Saddle, near the Seven Devils Wilderness, Idaho. Although the plants in that habitat aborted 59–69% of their flowers in 1998 and 1999, they aborted only 2–4% of flowers containing *Greya* eggs. Developed capsules were 20.5-fold more likely to have *Greya* eggs than aborted capsules in 1998 and 9.2-fold more likely in 1999 (both $\chi^2 P < 0.0001$).

Within four other habitats, *Greya* pollination and oviposition had no effect on floral development, indicating a commensal relationship in those populations (Table 2). A previous four-year

Table 1 Percentage of *Lithophragma parviflorum* plants with *Greya politella* eggs at sites within and near the northern Rocky Mountains, USA

River system	Plants with eggs (%)	Range over years	Total no. of plants sampled	No. of years sampled
Snake River, Washington	27	11–41	572	3
Turnbull NWR, Washington	93	85–98	599	3
Salmon River, Idaho				
Uplands	39	6–68	1,193	3
Lowlands	30	18–39	1,195	4
Clearwater River, Idaho	40	33–47	399	2
Selway River, Idaho	45	27–58	1,228	5
Rapid River, Idaho	66	36–49	348	4
Weiser River, Idaho	12	—	406	1
Snake River, Idaho	52	33–62	559	3
Wenaha River, Oregon	26	21–32	500	3

An average of 197 plants were sampled for each population in each year. All plants were separated by at least 1 m. Three capsules on floral positions 2–4 were dissected from each plant.

study using a different experimental protocol at one of these sites (Granite Point, Washington) demonstrated that abundant co-pollinators swamp the mutualistic effects of *G. politella* in that steppe habitat, reducing the outcome to commensalism^{20,21}. The earlier results (1987–90) together with the additional three years reported here a decade later (1997–99) indicate that the interaction at Granite Point remained commensal over multiple years.

The remaining four habitats showed evidence of an antagonistic relationship between the plants and the moths. Plants selectively aborted capsules containing *Greya* eggs, indicating current antagonistic coevolutionary hotspots (Table 2). Aborted capsules at these sites were on average 2.2-fold (range 1.6–2.4-fold between sites) more likely to have *Greya* eggs than developed capsules. For example, along the South Fork of the Clearwater River in Idaho, plants aborted 25% of flowers in both 1998 and 1999, but they aborted 34–38% of flowers containing *Greya* larvae.

These results indicate that the outcome of the interaction between *Greya* and *Lithophragma* differs significantly between habitats. We evaluated the outcome further by testing whether developing capsules visited by *Greya* had more or fewer seeds per flower than flowers not visited by the moths. Developing flowers with eggs did not differ significantly in mean seed number from flowers on the same plant without eggs at most sites and years. Only 4 of 33 comparisons between habitats and years (12 habitats for 2–3 years) showed significant differences in the mean number of seeds per flower, 2 in 1997 and 2 in 1999 (two-way analyses of variance or Wilcoxon signed-rank tests with $P < 0.05$ criterion; $N = 769,723$ seeds). Hence, the major effect of *Greya* is on whether capsules develop at all rather than on the number of seeds that mature within a capsule. Flowers not visited by *Greya* at the mutualistic sites often simply fail to develop through lack of pollination, whereas flowers at the antagonistic and commensal sites are sufficiently pollinated by co-pollinators for the positive effects of *Greya* on pollination to be swamped.

We quantified the relative rarity of co-pollinators at one of the mutualistic sites, Turnbull National Wildlife Refuge in Washington State, in 1999. *Greya* was the most common visitor, composing 42% of visits ($N = 1,146$ visits by all floral visitors). Another 32% of visits were by reliable co-pollinators, classified as reliable according to the experimental results of an earlier study²¹. The co-pollinators included several bombyliid fly and bee species^{20,21}. The remaining 26% were by insects whose visits rarely result in effective pollination, as determined by earlier studies²¹.

Visitation rate per flower was low. Averaged over the season, flowers had a probability of 0.47 of being visited by *Greya* on each of the one to several days for which they were open, but a probability of only 0.36 of being visited by a co-pollinator. Because only some visits by *Greya* and co-pollinators result in pollination²¹, these low visitation rates indicate that some flowers abscise through lack of pollination after the 1–2 days for which each flower remains open. Hence, pollination of flowers at Turnbull in 1999 relied strongly on visits by *Greya*, corroborating the results obtained in the capsule dissections (Table 2).

These overall results confirm a major prediction from recent coevolutionary theory: that geographically structured species will tend to coevolve toward a complex spatial mosaic of coevolutionary hotspots and coldspots. Although this is the most extensive analysis undertaken so far on the geographic structure of current phenotypic selection acting on species interactions across natural populations, it covers only part of the geographic range of this interspecific interaction. However, additional complex outcomes are likely to occur in other habitats, depending on the community context of the interaction. The results therefore indicate two general points about the geographic mosaic of coevolution and its effects on the organization of biodiversity. First, coevolution in geographically structured species may result in complex patterns of phenotypic selection among habitats. Recent work on other interspecific interactions has

shown differences in coevolved traits across broad geographic landscapes^{9–18}. The differences in current coevolutionary selection across even smaller-scale geographic landscapes in *Greya* and *Lithophragma* indicate that the long-term ebb and flow of coevolutionary selection might be important sources of ecological and evolutionary dynamics in natural biological communities. Second, if coevolving interactions commonly display selection mosaics and a mix of coevolutionary hotspots and coldspots, then severe anthropogenic fragmentation of landscapes is likely to have major effects on how continuing coevolution contributes to the persistence and dynamics of the Earth's biodiversity. □

Methods

Field surveys

Field surveys on the geographic occurrence of the interaction were conducted in the major river drainages of Idaho and adjacent Oregon and Washington by collecting the second, third and fourth capsules produced on plants distributed along linear transects running parallel to the river ($N = 6,999$ plants in total). All plants were separated by at least 1 m from all other plants. Capsules were placed in ethanol, returned to the laboratory and dissected under a microscope to determine the presence of *Greya* eggs or larvae.

Evaluation of outcome of the interaction

Although females also take nectar from the flowers, significant pollination by *Greya* occurs only during oviposition²⁵. Within each population we evaluated whether the second to the fourth flowers on each plant were more likely, or less likely, to develop mature seeds when they contained *Greya* eggs. Each plant generally produces four to eight flowers in sequence along a floral stalk. We chose the second to the fourth flowers because previous experiments had shown that plants rarely suffer resource limitation when maturing seeds within the first four flowers that they produce²¹. Hence, floral abortion through resource limitation is unlikely to affect these early flowers. The three-year analysis included the dissection of 18,912 floral capsules, including both developed capsules (that is, those with mature seeds) and aborted capsules (that is, those producing no mature seeds and withering on the plant).

We tested whether developed capsules visited by *Greya* had more or fewer seeds per flower than flowers not visited by the moths by comparing mean seed numbers for pairs of developed capsules: one with *Greya* eggs and one without (the mean number of floral pairs compared per habitat each year was 56.6; the total number of maturing seeds counted was 769,723). Analysis was restricted to randomly chosen pairs of second to fourth flowers taken from individual plants.

Quantification of pollinator visitation rates

We quantified pollinator visitation rates at Turnbull National Wildlife Refuge by recording all visits to *L. parviflorum* flowers within randomly chosen plots during 72 h of observation distributed across the peak week of the flowering season in 1999 ($N = 1,146$ visits by floral visitors). Each 1 m × 1 m plot contained multiple flowering individuals. Each observer recorded all visits to a plot within 1 h and then moved to a new plot.

Received 18 February; accepted 14 March 2002; doi:10.1038/nature00810.

1. Thompson, J. N. *The Coevolutionary Process* (Univ. Chicago Press, Chicago, 1994).
2. Hochberg, M. E. & van Baalen, M. Antagonistic coevolution over productivity gradients. *Am. Nat.* **152**, 620–634 (1998).
3. Nuismer, S. L., Thompson, J. N. & Gomulkiewicz, R. Gene flow and geographically structured coevolution. *Proc. R. Soc. Lond. B* **266**, 605–609 (1999).
4. Case, T. J. & Taper, M. L. Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. *Am. Nat.* **155**, 583–605 (2000).
5. Gomulkiewicz, R., Thompson, J. N., Holt, R. D., Nuismer, S. L. & Hochberg, M. E. Hot spots, cold spots, and the geographic mosaic theory of coevolution. *Am. Nat.* **156**, 156–174 (2000).
6. Hochberg, M. E., Gomulkiewicz, R., Holt, R. D. & Thompson, J. N. Weak sinks could cradle mutualisms—strong sources should harbour parasitic symbioses. *J. Evol. Biol.* **13**, 213–222 (2000).
7. Nuismer, S. L., Thompson, J. N. & Gomulkiewicz, R. Coevolutionary clines across selection mosaics. *Evolution* **54**, 1102–1115 (2000).
8. Holland, J. N. & Fleming, T. H. Geographic and population variation in pollinating seed-consuming interactions between senita cacti (*Lophocereus schottii*) and senita moths (*Upiga virescens*). *Oecologia* **121**, 405–410 (1999).
9. Berenbaum, M. & Zangerl, A. Chemical phenotype matching between a plant and its insect herbivore. *Proc. Natl Acad. Sci. USA* **95**, 13743–13748 (1998).
10. Benkman, C. W. The selection mosaic and diversifying coevolution between crossbills and lodgepole pine. *Am. Nat.* **153**, S75–S91 (1999).
11. Benkman, C. W., Holimon, W. C. & Smith, J. W. The influence of a competitor on the geographic mosaic of coevolution between crossbills and lodgepole pine. *Evolution* **55**, 282–294 (2001).
12. Brodie, E. D. III & Brodie, E. D. Jr Costs of exploiting poisonous prey: evolutionary trade-offs in a predator-prey arms race. *Evolution* **53**, 626–631 (1999).
13. Burdon, J. J. & Thrall, P. H. Spatial and temporal patterns in coevolving plant and pathogen associations. *Am. Nat.* **153**, S15–S33 (1999).
14. Kraaijeveld, A. R. & Godfray, H. C. J. Geographic patterns in the evolution of resistance and virulence in *Drosophila* and its parasitoids. *Am. Nat.* **153**, S61–S74 (1999).
15. Lively, C. M. Migration, virulence, and the geographic mosaic of adaptation by parasites. *Am. Nat.* **153**, S34–S47 (1999).
16. Radtkey, R. R., Fallon, S. M. & Case, T. J. Character displacement in some *Cnemidophorus* lizards revisited: A phylogenetic analysis. *Proc. Natl Acad. Sci. USA* **94**, 9740–9745 (1997).

17. Parker, M. A. Mutualism in metapopulations of legumes and rhizobia. *Am. Nat.* **153**, S48–S60 (1999).
18. De Jong, P. W., De Vos, H. & Nielsen, J. K. Demic structure and its relation with the distribution of an adaptive trait in Danish flea beetles. *Mol. Ecol.* **10**, 1323–1332 (2001).
19. Thompson, J. N. The evolution of species interactions. *Science* **284**, 2116–2118 (1999).
20. Thompson, J. N. & Pellmyr, O. Mutualism with pollinating seed parasites amid co-pollinators: constraints on specialization. *Ecology* **73**, 1780–1791 (1992).
21. Pellmyr, O. & Thompson, J. N. Sources of variation in pollinator contribution within a guild: the effects of plant and pollinator factors. *Oecologia* **107**, 595–604 (1996).
22. Caicco, S. L., Scott, J. M., Butterfield, B. & Csuti, B. A gap analysis of the management status of the vegetation of Idaho (U.S.A.). *Conserv. Biol.* **9**, 498–511 (1995).
23. Ricketts, T. H., Dinerstein, E., Olson, D. M. & Loucks, C. J. (eds) *Terrestrial Ecoregions of North America: A Conservation Assessment* (Island Press, Washington DC, 1999).
24. Pellmyr, O., Thompson, J. N., Brown, J. M. & Harrison, R. G. Evolution of pollination and mutualism in the yucca moth lineage. *Am. Nat.* **148**, 827–847 (1996).
25. Pellmyr, O. & Thompson, J. N. Multiple occurrences of mutualism in the yucca moth lineage. *Proc. Natl Acad. Sci. USA* **89**, 2927–2929 (1992).

Acknowledgements

We thank D. Althoff, E. Bartlett, R. Beck, N. Bonilla, S. Bradford, C. Clark, C. Cody, M. Jacoby, S. Lambert, Y. Martinez, K. Merg, S. Nuismer, J. Olmstead, S. Ringo, K. Segraves, D. Shepard, Z. Stanley and T. Steury for help in the field or laboratory, and R. Calsbeek, D. Hemby, K. Horjus, R. Hufft, N. Janz, K. Merg, S. Nuismer and J. Richardson for helpful discussions on the manuscript. We are grateful for permissions from the Clearwater, Nez Perce, Payette, Salmon, and Umatilla National Forests and from the Turnbull National Wildlife Refuge. This work was supported by the NSF.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.N.T. (e-mail: thompson@biology.ucsc.edu).

Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells

Kenneth D. Harris, Darrell A. Henze, Hajime Hirase, Xavier Leinekugel, George Dragoi, Andras Czurko & György Buzsáki

Center for Molecular and Behavioral Neuroscience, Rutgers, The State University of New Jersey, 197 University Avenue, Newark, New Jersey 07102, USA

According to the temporal coding hypothesis¹, neurons encode information by the exact timing of spikes. An example of temporal coding is the hippocampal phase precession phenomenon, in which the timing of pyramidal cell spikes relative to the theta rhythm shows a unidirectional forward precession during spatial behaviour^{2,3}. Here we show that phase precession occurs in both spatial and non-spatial behaviours. We found that spike phase correlated with instantaneous discharge rate, and precessed unidirectionally at high rates, regardless of behaviour. The spatial phase precession phenomenon is therefore a manifestation of a more fundamental principle governing the timing of pyramidal cell discharge. We suggest that intrinsic properties of pyramidal cells have a key role in determining spike times, and that the interplay between the magnitude of dendritic excitation and rhythmic inhibition of the somatic region is responsible for the phase assignment of spikes^{4,5}.

Two competing frameworks have been suggested for the representation of information in the brain: rate coding and temporal coding. Rate coding suggests that information is represented by discharge rates over a group of neurons⁶. Temporal coding models, on the other hand, argue that the temporal relationship of spikes is more critical for information transfer^{7–11}. During spatial behaviours, hippocampal pyramidal neurons discharge specifically at

certain locations in the environment (the 'place field'), implying that hippocampal neurons 'code' for spatial information¹². Furthermore, phase assignment of spikes with respect to the theta rhythm shows a unidirectional advancement as the animal crosses the place field^{2,3}. Because spike phase shows a monotonic relation to position, whereas firing rate increases and decreases as the animal enters and leaves the place field, it was suggested that this 'temporal code' might be a reliable mechanism for representing space^{2,13}.

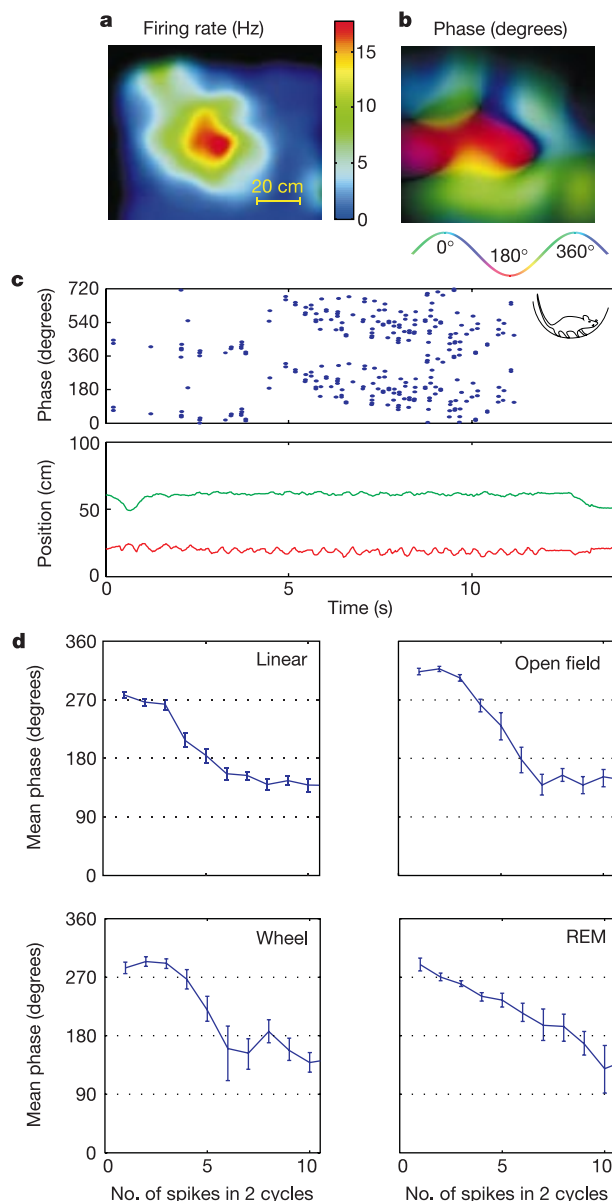


Figure 1 Relation between theta phase and instantaneous firing rate in pyramidal cells. **a**, Mean firing rate as a function of position, for an illustrative CA1 pyramidal cell ('place map'). **b**, Mean firing phase as a function of position for the same cell ('phase map'). **c**, Spike phase advancement during a wheel running episode. The cell fired sparsely at constant phase until 5 s into the trial, when an intense period of firing occurred, accompanied by phase precession, although the rat's head remained stationary (<3 cm). Top panel, phase and time of spikes. Bottom panel, instantaneous head position (red and green: X and Y coordinates). **d**, Rate versus phase relationship during four types of behaviour. For each spike, instantaneous rate was estimated by the number of spikes in the 720° interval centred on it. Graph shows circular mean and 95% confidence interval²³ of all spikes in the database, grouped by instantaneous rate. Phase 0 corresponds to the positive peak of CA1 pyramidal layer theta. The database consisted of 28, 181, 23 and 253 cells from 3, 5, 7 and 15 rats in the linear, open field, wheel and REM behaviours.

A critical issue is whether phase precession is explicitly tied to spatial computations, or arises from a more general mechanism. We suggest that hippocampal phase advancement has a cellular origin, owing to a dependence of preferred phase on the strength of dendritic depolarization^{4,5}. If this is the case, phase precession effects should be seen in both spatial and non-spatial behaviours. Here we address this issue by investigating the relation of phase to spike train dynamics, independent of continuing behaviour.

Figure 1a and b shows the two-dimensional place field and 'phase field' of a representative pyramidal cell during spatial exploration in an open field, showing the firing rate and mean phase of firing with respect to CA1 pyramidal layer theta at each point in space. In agreement with previous investigations^{2,3}, both firing rate and phase correlated with the spatial coordinates of the animal. In agreement with the depolarization hypothesis, the spatial dependence of firing phase largely mirrored that of firing rate, with mean phase in the centre of the place field being advanced compared to the periphery. An example of phase precession during a non-spatial behaviour (wheel running) is shown in Fig. 1c, where again phase precession was seen at times of high firing rate.

To investigate the phase–firing rate relationship quantitatively and independently of the spatial position of the rat (Fig. 1d), we examined the phase of spikes during linear track running ($n = 28$ cells), open field exploration ($n = 181$), wheel running with stationary head ($n = 23$), and REM sleep ($n = 253$). Instantaneous firing rate was quantified for each spike by counting the number of spikes up to 360° of theta before and after the spike whose rate was to be quantified. For each behavioural category, the effect of firing rate on phase was significant ($P < 0.001$; circular analysis of

variance (ANOVA) on all spikes for each behaviour; see Supplementary Information). Mean phase, averaged over all behaviours, shifted from $295^\circ \pm 7^\circ$ (circular mean $\pm$ 95% confidence interval) at low instantaneous rates (1 spike per 2 cycles; ~ 4 Hz) to $141^\circ \pm 9^\circ$ for high rates (≥ 10 spikes per 2 cycles; ~ 40 Hz). CA3 pyramidal cells recorded in linear track running ($n = 54$) and open field ($n = 74$) showed a similar relationship, although the phase of firing (relative to CA1 pyramidal layer theta) was earlier at both low and high firing rates (overall mean $258^\circ \pm 15^\circ$ and $96^\circ \pm 25^\circ$). As a further test of the proposed depolarization–phase relationship, we examined the relationship of spike phase on extracellular amplitude, which has been observed to decrease at times of strong activity^{14,15}. In each behavioural category, spike phase advanced robustly as the amplitude decreased (Supplementary Information).

A phase–firing rate correlation therefore exists in both spatial and non-spatial behaviours. Nevertheless, this correlation may not fully summarize the phase dynamics of pyramidal cells. In spatial behaviours, phase shift occurs unidirectionally over multiple theta cycles^{2,3}. Do spike trains in non-spatial behaviours show a similar tendency? Unidirectional phase advancement implies that the phase of spikes advances between the onset and the cessation ('offset') of periods of firing. On a linear track, firing onset and offset reliably occur in distinct spatial locations, and unidirectional advancement may therefore be detected by correlating phase with position. However, this analysis cannot be used in other behaviours. We therefore classified spikes as corresponding to firing onset and offset directly. Spikes were classified as belonging to an accelerating spike

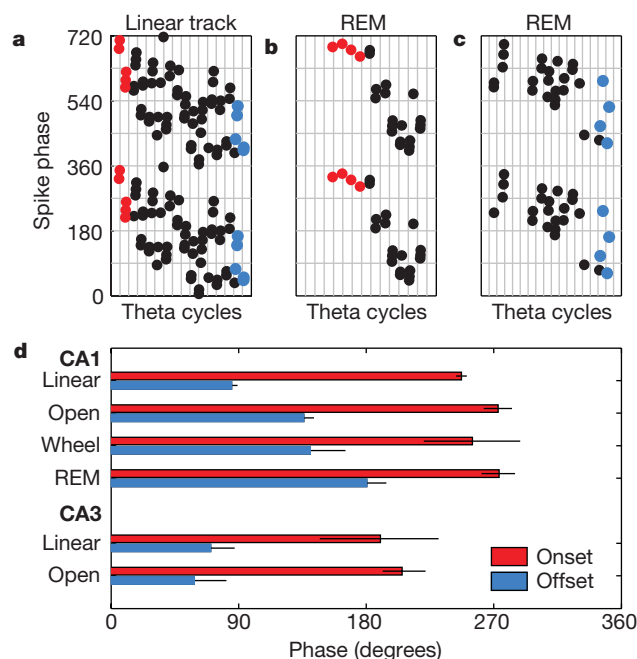


Figure 2 Unidirectional nature of the phase shift. **a–c**, Subsets of spikes were selected as occurring at the onset (red) or offset (blue) of activity (see text). Most spikes (black) met neither criterion. The graphs show the results of the selection process on example spike trains from three single units. Out of a grand total of 301,693 spikes, 7,750 were selected as onset and 7,277 as offset. Vertical lines, successive theta cycles. **d**, For each behaviour, the circular mean phase and 95% confidence interval was calculated for all spikes in the database classified as onset (red) or offset (blue). For CA1, number of cells and rats same as Fig. 1. For CA3, 54 and 74 cells from 1 rat in the linear track and open field behaviour.

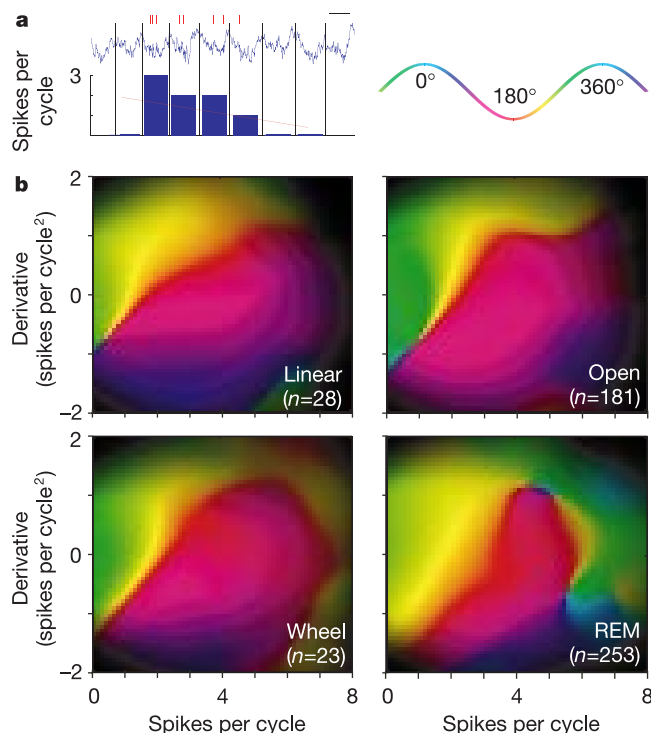


Figure 3 Mean phase as a function of instantaneous rate and its temporal derivative. **a**, Calculation of rate derivative. Top, theta oscillation (blue) and spike times (red ticks) in an example data segment. Scale bar, 100 ms. Bottom, the temporal derivative of discharge rate for each cycle was calculated by linear regression of spike count (blue bars) against cycle number in a seven-cycle window (red line). **b**, Pseudo-colour plots showing phase as a function of instantaneous rate and its derivative, averaged over all spikes in the database, for each behaviour. Rate derivative begins to influence spike phase at rates > 1 spike per theta cycle (~ 8 Hz). Number of cells and rats same as Fig. 1.

train (onset) if the previous 8 theta cycles contained at most 4 spikes, and the following 8 contained at least 16 (an increase from <4 Hz to >16 Hz, assuming 8-Hz theta), and to a decelerating train (offset) if the previous 8 cycles contained at least 16, and the next at most 4 (Fig. 2a–c). The mean phase was significantly earlier for spikes in decelerating patterns than for accelerating patterns in all four behaviours (Fig. 2d; $P < 0.001$; Mardia-Watson-Wheeler test; Supplementary Information).

Is the unidirectional nature of phase precession explained simply by lower firing rates at activity onset, or can it occur even in the absence of firing rate change? To address this question, we considered the dependence of phase on both rate and its instantaneous temporal derivative (Fig. 3a, b). At low firing rates (≤ 0.25 spikes per cycle, ~ 2 Hz), there was little effect of rate derivative (mean of all spikes in combined behaviours $\pm 95\%$ confidence interval: $317^\circ \pm 12^\circ$; green on Fig. 3b). At higher rates (>1 spike per cycle, ~ 8 Hz), mean spike phase depended on the combined effect of rate and the derivative of rate, advancing further as acceleration (mean of all spikes with derivative ≥ 1 spike per cycle², $267^\circ \pm 28^\circ$; yellow),

turned to deceleration (mean of all spikes with derivative <1 spike per cycle², $92^\circ \pm 26^\circ$; mauve). The example of Fig. 1c is therefore illustrative of mean behaviour: at low rates, phase is constant, and unidirectional advancement occurs only at times of intense activity¹⁶.

The finding that phase advancement occurs only at times of high firing rate supports the hypothesis that phase precession is a function of excitation strength. According to this hypothesis, phase modulation occurs due to an ‘interference’ effect between rhythmic somatic inhibition and dendritic excitation (soma-dendritic interference or SDI model)^{4,5}. At times of weak drive, inhibition dominates, and the cell fires when it is least inhibited (the positive phase). At times of strong drive, excitation dominates, and the cell fires when it is most excited (the negative phase). This interference mechanism may also explain the unidirectional nature of phase precession, given a further assumption that, owing to adaptation, the firing rate begins to decrease before the strength of dendritic excitation has reached its peak. Under this assumption, the onset of spiking activity will occur at the same time as the onset of excitation, but the offset of spiking will occur before the offset of excitation. Accordingly, dendritic depolarization will be larger at the offset of spiking than at the onset, and phase will be advanced.

This hypothesis was investigated in two ways. First, temporally symmetric current ramps were injected into CA1 and CA3 pyramidal cells of anaesthetized rats ($n = 18$; Fig. 4a, b). The resulting spike trains showed 2.6 ± 0.7 times more spikes on the ascending than the descending part of the ramp, supporting the hypothesis that firing rate begins to decline before the offset of excitation. Furthermore, the degree of asymmetry increased with increasing amount of depolarization. Second, a compartmental model pyramidal cell (Fig. 4c; see Methods) was subjected to a sinusoidal somatic Cl^- conductance together with a sinusoidal dendritic Na^+ conductance. The latter was modulated by a symmetrical ramp function, simulating a place field traversal. The model exhibited a dependence of preferred phase on excitation strength. In addition, because of adaptation, firing offset occurred before excitation offset, and a unidirectional phase advance was seen, similar to place field traversals *in vivo*^{2,3}.

In summary, in both spatial and non-spatial behaviours, timing of hippocampal pyramidal cells within the theta cycle shows a correlation with instantaneous firing rate, and exhibits unidirectional phase precession at rates exceeding one spike per cycle. Similar phase precession was produced by a SDI model, in which phase was determined by the relative magnitude of oscillatory inhibitory somatic and excitatory dendritic inputs. We propose that the SDI mechanism has an important role in coordinating the timing of hippocampal cell populations, and that this temporal organization is fundamental to processing of both spatial and non-spatial information in the hippocampus. □

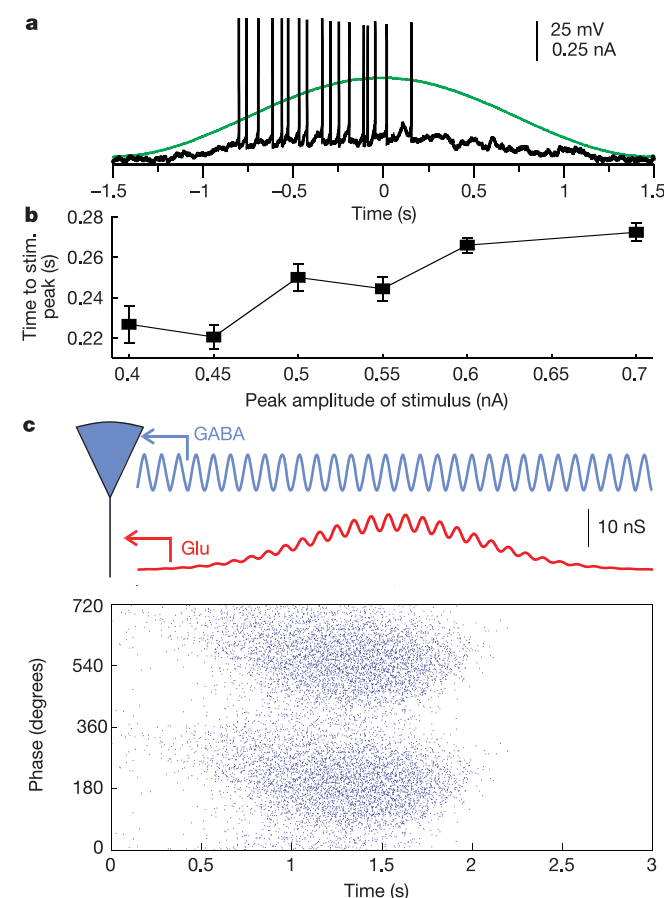


Figure 4 Proposed role of adaptation in unidirectional phase precession. **a**, Symmetrical ramp currents (green line), were injected into the soma of hippocampal pyramidal neurons of anaesthetized rats. More spikes occurred on the ascending than descending part of the ramp, and peak firing preceded the depolarization peak. **b**, The forward shift between peak firing rate and maximum depolarization increased with the amount of depolarization. Each point shows mean and standard error of 50 repetitions from one cell. **c**, Soma-dendritic interference (SDI) model of phase precession. A compartmental model pyramidal cell experienced a sinusoidal Cl^- conductance at the soma (mimicking GABA_A receptor-mediated inhibition), and a sinusoidal dendritic Na^+ conductance (mimicking glutamate excitation, Glu), modulated by a symmetrical drive function. The graph shows phase versus time for the overlaid spikes of 100 model runs. Owing to adaptation, the model ceased firing before the offset of excitation, and showed unidirectional phase advance.

Methods

Fifteen male rats were implanted with movable tetrodes. Before implantation, seven rats were trained to run continuously in a running wheel for water reinforcement available in an adjacent box. The remaining eight animals were recorded either while exploring in a large rectangular box ($1.2 \text{ m} \times 1.2 \text{ m} \times 0.5 \text{ m}$ high) or walking on an elevated square track (65 cm side length; 8.5 cm wide ‘linear’ track) for chocolate reward. An infrared LED was attached to the head stage to track the position of the animal. Sleep recordings were carried out in the home cage of the rat. To ensure stationarity, the first and last seconds of wheel running epochs were excluded. Electrode placement was localized histologically¹⁷. Extracellular spikes were extracted from the traces by previously described methods^{18,19}. Only sufficiently well-isolated units were considered for analysis (isolation distance ≥ 20)¹⁵. Theta activity was detected by calculating the ratio of the Fourier components of the theta ($5\text{--}10$ Hz) and delta ($2\text{--}4$ Hz) frequency band²⁰. Intracellular experiments were performed in urethane anaesthetized rats²¹, with intracellular injection of symmetrical triangular or sinusoid waveforms (3 s long).

Local field potentials recorded in the CA1 pyramidal layer were digitally filtered for theta ($4\text{--}10$ Hz). Spike phase was calculated using a Hilbert transform. ‘Phase fields’ were calculated by a locally weighted circular mean using a gaussian smoothing function of width 7.5 cm . To compute the derivative of firing rate, a linear regression was performed for the number of spikes per cycle as a function of cycle number, using the central and three

neighbouring cycles on each side. The derivative and instantaneous firing rate were estimated as the slope of the fit line, and its intercept in the central cycle. Circular ANOVA was performed using the test statistic $\sum \cos(\theta_i - \bar{\theta}_{g(i)})$, where $\bar{\theta}_{g(i)}$ is the circular mean for the group to which observation i is assigned. The null distribution was computed by 1,000-fold random reassignment of groups.

Computational simulations were performed in NEURON (<http://www.neuron.yale.edu>) using a two-compartment model of a hippocampal pyramidal cell. Conductance and calcium dynamics were as described by Migliore²². Compartment geometry: soma length 50 μm , diameter 40 μm , dendrite length 50 μm , diameter 3.3 μm . Somatic peak channel conductances (S cm^{-2}): G_{Na} 0.015; $G_{\text{K}}(\text{DR})$ (delayed rectifier) 0.009; $G_{\text{K}}(\text{A})$ (A-type) 0.0001; $G_{\text{K}}(\text{M})$ (M-type) 0.00002; $G_{\text{K}}(\text{AHP})$ (after-hyperpolarization) 0.004; $G_{\text{Ca}}(\text{L})$ (L-type) 0.0025; $G_{\text{Ca}}(\text{N})$ (N-type) 0.0025; $G_{\text{Ca}}(\text{T})$ (T-type) 0.00025; dendritic same without G_{Na} . 10-Hz sinusoidal conductances were added to the soma (Cl^- , peak 20 nS at 120°, phase modulation 50%) and dendrite (Na^+ , peak 15 nS at 180°, phase modulation 30%). Modulation depth and preferred phase were based on extracellular recordings of interneurons and pyramidal cells *in vivo*²⁰. The dendritic current was further modulated by a gaussian function of time (width 500 ms). Synaptic noise was added at the dendrite (Na^+ , Poisson process, mean rate 1 kHz, peak conductance 5 nS).

Received 8 November 2001; accepted 2 April 2002; doi:10.1038/nature00808.

1. Singer, W. Synchronization of cortical activity and its putative role in information processing and learning. *Annu. Rev. Physiol.* **55**, 349–374 (1993).
2. O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
3. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
4. Magee, J. C. Dendritic mechanisms of phase precession in hippocampal CA1 pyramidal neurons. *J. Neurophysiol.* **86**, 528–532 (2001).
5. Kamondi, A., Acsády, L., Wang, X. J. & Buzsáki, G. Theta oscillations in somata and dendrites of hippocampal pyramidal cells in vivo: activity-dependent phase-precession of action potentials. *Hippocampus* **8**, 244–261 (1998).
6. Adrian, E. D. & Zotterman, Y. The impulses produced by sensory nerve endings. Part 2. The response of a single end organ. *J. Physiol. (Lond.)* **61**, 151–171 (1926).
7. Gray, C. M. & Singer, W. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl Acad. Sci. USA* **86**, 1698–1702 (1989).
8. Hopfield, J. J. Pattern recognition computation using action potential timing for stimulus representation. *Nature* **376**, 33–36 (1995).
9. Buzsáki, G. & Chrobak, J. J. Temporal structure in spatially organized neuronal ensembles: a role for interneuronal networks. *Curr. Opin. Neurobiol.* **5**, 504–510 (1995).
10. Vaadia, E. *et al.* Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* **373**, 515–518 (1995).
11. Riehle, A., Grun, S., Diesmann, M. & Aertsen, A. Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* **278**, 1950–1953 (1997).
12. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, 1978).
13. Burgess, N., Recce, M. & O'Keefe, J. A model of hippocampal function. *Neural Networks* **7**, 1065–1081 (1994).
14. Quirk, M. C., Blum, K. I. & Wilson, M. A. Experience-dependent changes in extracellular spike amplitude may reflect regulation of dendritic action potential back-propagation in rat hippocampal pyramidal cells. *J. Neurosci.* **21**, 240–248 (2001).
15. Harris, K. D., Hirase, H., Leinekugel, X., Henze, D. A. & Buzsáki, G. Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* **32**, 141–149 (2001).
16. Hirase, H., Czurko, A., Csicsvari, J. & Buzsáki, G. Firing rate and theta-phase coding by hippocampal pyramidal neurons during 'space clamping'. *Eur. J. Neurosci.* **11**, 4373–4380 (1999).
17. Csicsvari, J., Hirase, H., Mamiya, A. & Buzsáki, G. Ensemble patterns of hippocampal CA3–CA1 neurons during sharp wave-associated population events. *Neuron* **28**, 585–594 (2000).
18. Csicsvari, J., Hirase, H., Czurko, A. & Buzsáki, G. Reliability and state dependence of pyramidal cell-interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. *Neuron* **21**, 179–189 (1998).
19. Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H. & Buzsáki, G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophysiol.* **84**, 401–414 (2000).
20. Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A. & Buzsáki, G. Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *J. Neurosci.* **19**, 274–287 (1999).
21. Henze, D. A. & Buzsáki, G. Action potential threshold of hippocampal pyramidal cells in vivo is increased by recent spiking activity. *Neuroscience* **105**, 121–130 (2001).
22. Migliore, M., Cook, E. P., Jaffe, D. B., Turner, D. A. & Johnston, D. Computer simulations of morphologically reconstructed CA3 hippocampal neurons. *J. Neurophysiol.* **73**, 1157–1168 (1995).
23. Fisher, N. I. *Statistical Analysis of Circular Data* (Cambridge Univ. Press, New York, 1993).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

This work was supported by the NIH and the Human Frontier Science Program.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.B. (e-mail: buzsaki@axon.rutgers.edu).

Role of experience and oscillations in transforming a rate code into a temporal code

M. R. Mehta, A. K. Lee & M. A. Wilson

Center for Learning & Memory, Department of Brain & Cognitive Sciences, RIKEN-MIT Neuroscience Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

In the vast majority of brain areas, the firing rates of neurons, averaged over several hundred milliseconds to several seconds, can be strongly modulated by, and provide accurate information about, properties of their inputs. This is referred to as the rate code. However, the biophysical laws of synaptic plasticity require precise timing of spikes over short timescales (<10 ms)^{1,2}. Hence it is critical to understand the physiological mechanisms that can generate precise spike timing *in vivo*, and the relationship between such a temporal code and a rate code. Here we propose a mechanism by which a temporal code can be generated through an interaction between an asymmetric rate code and oscillatory inhibition. Consistent with the predictions of our model, the rate^{3,4} and temporal^{5–7} codes of hippocampal pyramidal neurons are highly correlated. Furthermore, the temporal code becomes more robust with experience. The resulting spike timing satisfies the temporal order constraints of hebbian learning. Thus, oscillations and receptive field asymmetry may have a critical role in temporal sequence learning.

We recorded the activity of pyramidal neurons from the dorsal CA1 region of the hippocampus in awake, behaving rats (see Methods). The firing rate of these neurons depends on the spatial location of the rat³, and hence these neurons are referred to as 'place cells'. The mean firing rates (averaged over >200 ms) of about 100 neurons can provide an accurate estimate of the rat's spatial location⁴. This is the hippocampal rate code.

In addition to this spatial parameter, hippocampal activity during active exploration is strongly modulated by a temporal parameter, namely the 8-Hz theta rhythm. In a classic study, O'Keefe and Recce showed that^{5,6} the phase of the theta rhythm at which a place cell fires a spike steadily precesses to lower values as the rat moves further along the place field (Fig. 1a). Consistent with this, phase was highly (negatively) correlated with position ($r = -0.50 \pm 0.01$, $P < 0.0001$) across 171 recorded place fields (henceforth referred to as the 'population'). Normalizing by occupancy (see Methods) yields the firing rate as a function of position and phase—that is, the spatio-temporal receptive field (STRF, Fig. 1b). The firing rate increases as phase decreases (from 360° to 180°) and position increases, reaching a maximal value at a position beyond the centre (white line) of the place field and at 200° (red arrow) towards the end of the place field: that is, the firing rate is an inseparable function of high phases (180° < phase < 360°) and position. The firing rate at low phases (0° < phase < 180°) shows a weaker dependence on phase and position.

The vast majority of neurons have a similar spatio-temporal structure (as Fig. 1b), in which the firing rate shows a significant increase as the mean phase decreases and the distance within the place field increases (population average, Fig. 1c, d). Thus, there is spatial information in the precise timing of spikes (accurate up to 10 ms) with respect to the theta rhythm^{5–7}. This is the hippocampal temporal code, which has several advantages over the rate code, such as insensitivity to the rat's running speed⁵ and scale invariance⁷.

Several computational models have been proposed to explain the origin of this temporal code^{8–11}. These models were based upon the

observation that whereas the place field firing rate is a symmetric function of position, phase is not, suggesting that phase and rate are governed by separate mechanisms. However, recent results have demonstrated that place fields become asymmetric in an experience-dependent fashion¹², raising the possibility that rate and phase codes may be more directly coupled. Place field firing rate distributions in the present study were also asymmetric (skewness, -0.42 ± 0.05 ; $P < 0.0001$), such that the firing rate was low at the beginning of a place field, but was high at the end. On the basis of this observation we sought to test a hypothesized mechanism of the origin of the temporal code (Fig. 2), in which CA1 neurons receive theta rhythmic inhibitory input¹³, as well as asymmetric excitatory input^{12,14,15}. Under this model, the relationship between phase and rate is established by assuming that at the beginning of a place field excitation is low, and hence the latency with which a neuron comes out of inhibition and fires a spike, that is, spike phase, is high (360°). As the rat moves further along the place field, excitation increases, resulting in shorter latency to spiking (smaller phases), and higher rates. Such a mechanism, when applied to groups of overlapping place cells, will faithfully reproduce the temporal order of activation of place cells on short (<10 ms) timescales.

This mechanism makes the basic prediction that regions of high firing rate should have a lower mean phase, and vice versa. Indeed, mean phase was negatively correlated with mean firing rate both for individual cells (Fig. 3a, b) and across the population of neurons (Fig. 3c).

A second prediction of the model is that as the excitatory input increases with distance along the place field, the fraction of theta cycle during which excitation exceeds inhibition will also increase (blue bars, Fig. 2a), resulting in a wider phase distribution. Consistent with this prediction, we found an increase in the width of the phase distribution as a function of both distance (Fig. 3a, d) and rate ($r = 0.13 \pm 0.02$, $P < 0.0001$). Thus the temporal code was highly asymmetric, becoming less precise as the rat's position in the place field increased.

Further, excitation would be an increasing function of position in the first half of the place field, but initially increasing and then slightly decreasing function of position in the last half (Fig. 2). This predicts that the phase will be a more consistently decreasing function of position in the first half of the place field, compared to the last half. Thus the temporal code should be more robust in the first half of the place field than in the last half. Consistent with this, for the cell in Fig. 3a the phase is 6.2 times more strongly correlated with distance for spikes in the first half of the place field, compared with spikes in the last half of the place field. Similar results were true for the population of cells (Fig. 3e).

If the theta rhythm is purely sinusoidal, increasing levels of excitation will result in phase advancement from 360° at the trough of the theta cycle to 180° at the peak of the theta cycle. Beyond this point, the excitation exceeds inhibition throughout the theta cycle, resulting in a broader distribution of spikes. Thus, according to this model, the high phases should be more strongly correlated with

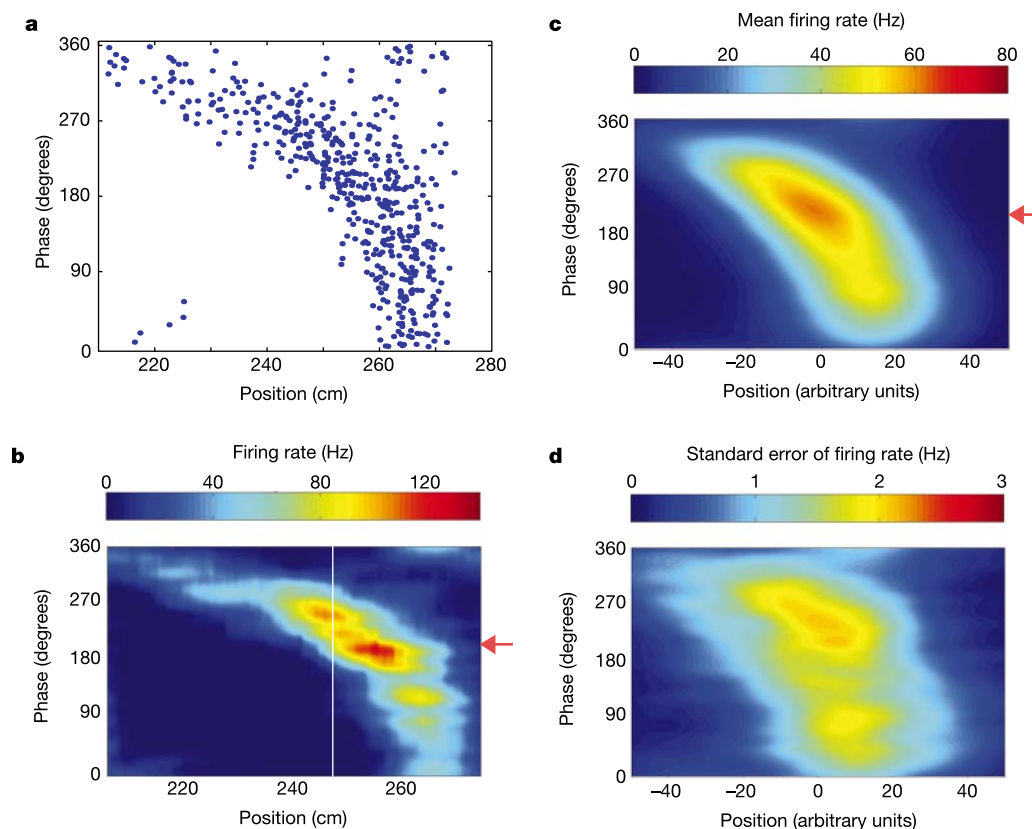


Figure 1 Hippocampal spatio-temporal receptive fields. **a**, The rat ran from left to right. Each point represents the phase of the theta rhythm (see Methods) and the rat's position at the time of occurrence of a spike from an isolated CA1 pyramidal neuron. Phase is highly correlated with position ($r = -0.8$). **b**, The firing rate for the neuron in **a** plotted as a function of position and phase (the spatio-temporal receptive field, STRF; see Methods). The centre of the place field (247 cm) is indicated by a white line. Phase (200°) corresponding to maximal firing rate is indicated by a red arrow. **c**, Population-averaged

STRF. Each place field was re-scaled to be 100 pixels (66 cm) wide. The centres of the re-scaled STRFs were aligned at the origin. The average value of firing rate in each spatio-temporal bin is plotted. Phase (220°) corresponding to maximal firing rate is indicated by a red arrow. **d**, The standard error of firing rate was computed for the re-scaled population-averaged STRF (**c**), and plotted as a function of phase and position, showing that the firing rate modulations in **c** are highly significant.

position than the low phases. Indeed, the spikes at high phases for the cell in Fig. 3a are 7.4 times more strongly correlated with position than spikes at low phases. As a population, the correlation with position was 2.5 times stronger for high phases than low phases (Fig. 3f). The residual precession at low phases could arise because the theta rhythm is not purely sinusoidal¹⁶.

The above results describe phase precession under the condition of an asymmetric receptive field. However, previous work has shown that place fields do not have a clear asymmetry during early exposure to an environment (Fig. 2b)¹². In this condition, the rate will not be an increasing function of space, and hence phase will be poorly correlated with position. Thus, if the phase was indeed determined by the above mechanism, the temporal code should be less robust during this early period of exposure (Fig. 4a). Consistent with this prediction, we found that on average the population of place cells had a 2.1 times stronger correlation of phase with position after experience, compared to before (Fig. 4b, $P < 0.0001$). The time course of this evolution of the temporal code was also comparable to the time course of evolution of place field firing rate asymmetry¹².

In order to infer the precise effect of experience on firing rate as a function of position and phase, we computed the population-averaged STRF (Fig. 1c) for each lap. The resulting STRF shows a significant 2.2 times stronger correlation ($P < 0.00001$, paired t -test) of phase with position later during experience (sixteenth lap, Fig. 4d) than earlier (first lap, Fig. 4c). Whereas the correlation between firing rate and position, and between phase and position, becomes stronger with experience, the model predicts that there should be no change in the correlation between rate and phase. Consistent with this, the rate and phase were significantly correlated in both the first ($r = -0.10 \pm 0.03$, $P < 0.01$) and the sixteenth ($r = -0.07 \pm 0.03$, $P < 0.01$) laps, and there was no significant change in this correlation with experience ($P > 0.8$).

In three out of seven sessions, the rat's running velocity showed a significant change with experience (see Supplementary Information). Hence, we restricted the analysis to 67 place fields obtained from the remainder of four sessions where there was no significant change in the rat's running velocity throughout the first seventeen laps. The magnitude and time course of the experience dependence of the temporal code in these restricted data were virtually identical (see Supplementary Information) with the results presented in Fig. 4b. Thus the lap-specific changes in the temporal code could not arise from changes in the rat's behaviour.

Finally, if the phase precession is indeed restricted to mostly high phase spikes and is a result of the asymmetric nature of the firing rate distribution, the lap-by-lap fluctuation in these two parameters should be correlated. Indeed, averaged across the population, the firing rate asymmetry in a given lap was highly significantly correlated with the correlation of high phase spikes with position ($r = 0.13 \pm 0.02$, $P < 0.0001$).

Thus, our model suggests that the phase of spikes at a given location is largely determined by the net excitatory input at that location: a CA1 pyramidal neuron fires a spike when feed-forward excitation exceeds periodic inhibition. This simple mechanism would be modified in the presence of recurrent inhibition as follows. A pyramidal neuronal spike would activate recurrent inhibition within CA1, resulting in a suppression of subsequent pyramidal neuronal activity in that theta cycle. The suppression would have two complementary effects on the rate and the temporal codes. First, the pyramidal activity would be restricted to a small part of the theta cycle, thereby making the temporal code sharper. Second, the suppression would be strongest at higher rates, found towards the end of the place field. Therefore, the firing rate asymmetry would provide an underestimate of the asymmetry of the excitatory drive on a pyramidal neuron. Hence, the spike phase would become a better estimator than spike rate of the strength of feed-forward

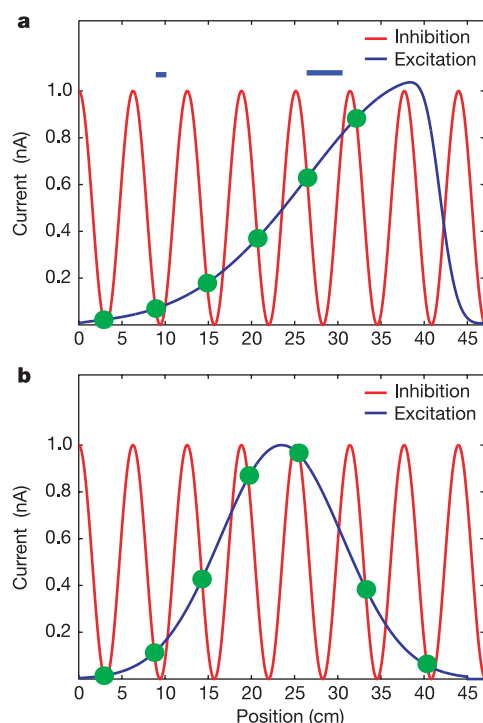


Figure 2 A mechanism that can generate a temporal code from a rate code. **a**, Each CA1 cell receives an asymmetric excitatory input (blue curve¹²) and periodic inhibitory input (red curve). The minima of inhibition correspond to 0° or 360°, and the maxima to 180°. The neuron will commence firing only when the excitation exceeds inhibition (green dots). Recurrent inhibition would then quickly terminate the activity. As the rat moves deeper into

the place field, the excitation increases, phase decreases, and the neuron is active for a larger fraction of the theta cycle (blue bars) **b**, Before experience, the excitatory inputs to the CA1 neuron do not have a significant asymmetry (blue curve¹², this does not necessarily mean that individual place fields are symmetric). Therefore, the firing rate, and hence the phase, will be poorly correlated with position.

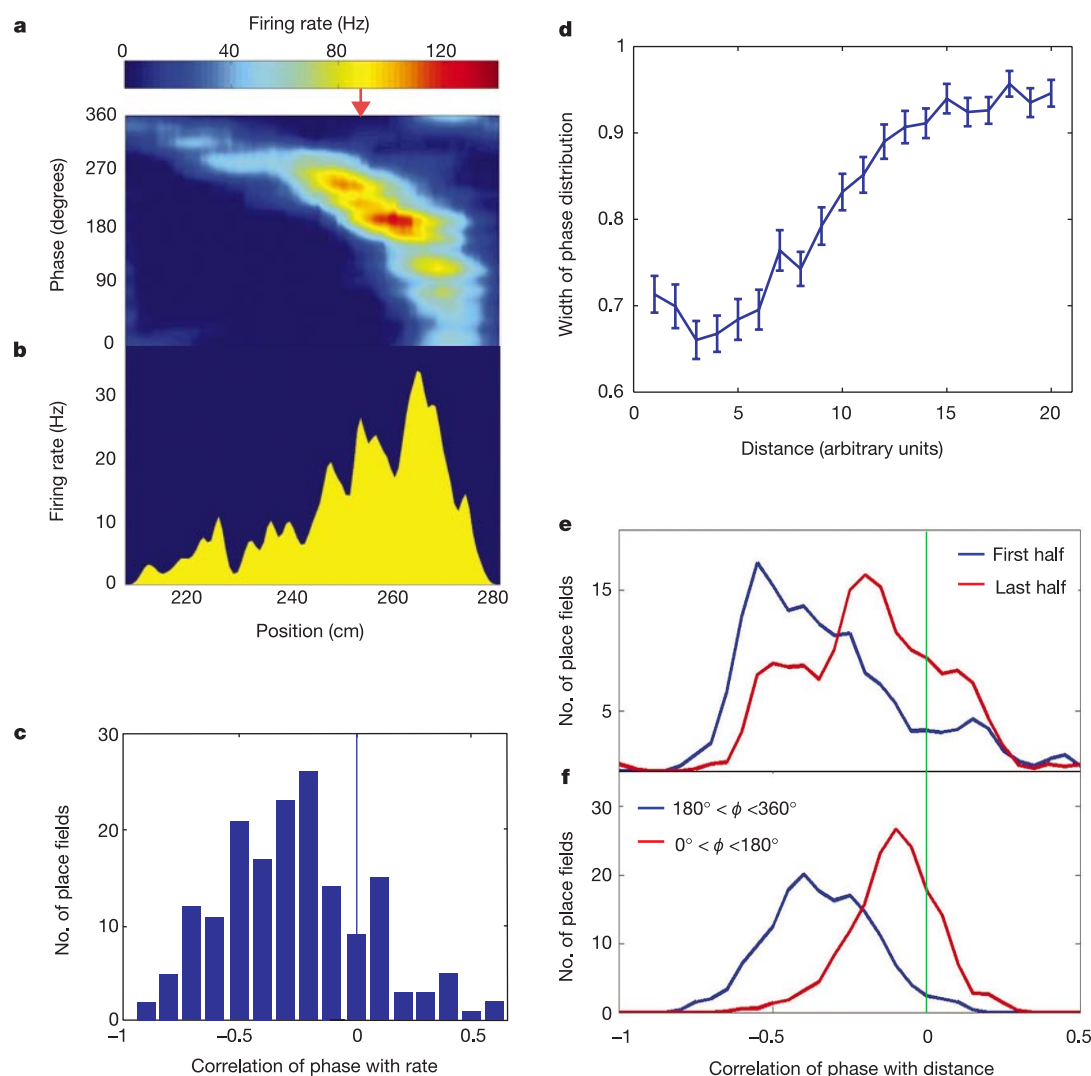


Figure 3 Relationship between hippocampal rate and temporal codes. The mean phase at a spatial location (**a**) is highly correlated with the mean firing rate (**b**) at that spatial location ($r = -0.6$). The place field centre is indicated by a red arrow. **c**, A histogram of correlation between rate and phase (see Methods) for the population of 171 cells shows that phase is negatively correlated with rate ($r = -0.30 \pm 0.02$, $P < 0.0001$). **d**, The width of phase distribution is larger at the end than at the beginning of the place field. Population-averaged value of the correlation of width of the phase distribution with

position was $r = 0.50 \pm 0.03$, $P < 0.0001$. **e**, Histograms of correlation of phase with position in the first half of the place field (blue line, $r = -0.33 \pm 0.02$) is 1.7 times stronger than that in the last half (red line, $r = -0.20 \pm 0.02$, $P < 0.0001$). **f**, Histograms of correlation of spikes at high phases ($\phi = 180^\circ$ – 360° , blue line) and low phases ($\phi = 0^\circ$ – 180° , red line). The high phases are 2.5 times more correlated with position ($r = -0.34 \pm 0.01$) than the low phases ($r = -0.13 \pm 0.01$, $P < 0.0001$).

excitation and hence of the rat's spatial location. This could explain the observed stronger correlation between phase and position than between phase and rate, and the stronger correlation between phase and position than between rate and position. This model does not explicitly incorporate other elements of the hippocampal system, and hence it does not rule out the potential contribution of other sequence-retrieval-based models of phase precession^{8,9,17,18}.

Although previous experimental studies of place fields on linear tracks^{5,6} did not explicitly investigate a relationship between the rate and phase, our results are in broad agreement with previous work¹⁹ on a non-spatial wheel running task where the mean firing rate of a cell was negatively correlated with the mean phase of that cell. Recent experiments have suggested that NMDA (*N*-methyl-D-aspartate) antagonists do not eliminate phase precession. These experiments showed an increased rate of phase precession²⁰ with NMDA antagonists, which is consistent with our data (larger rate of precession before experience than after). The lack of direct measure-

ment of the asymmetry of individual place fields and the correlation of phase with position and with rate in that study prevents direct comparison of their data with our model.

Recent *in vitro* experiments have shown that increasing amounts of current injection, coupled with theta-like oscillations, indeed result in phase advancement²¹. Consistent with our results^{12,14}, these *in vitro* experiments also obtained a clear relationship between the amplitude of injected current and phase in only a restricted part of the phase space. A similar relationship between rate and latency has also been observed in the STRF of direction selective neurons in V1 (see ref. 22 for a recent review), suggesting that similar mechanisms may be involved^{12,14,15,23,24}. The phase coding of orientation selectivity²⁵ can also be explained by a similar mechanism. An optimally oriented bar would excite a cell maximally, resulting in spiking at the peaks of the gamma rhythm, whereas a sub-optimally tuned bar would correspond to lesser excitatory drive, resulting in a phase lag.

In its most general form, our model suggests that when a stimulus

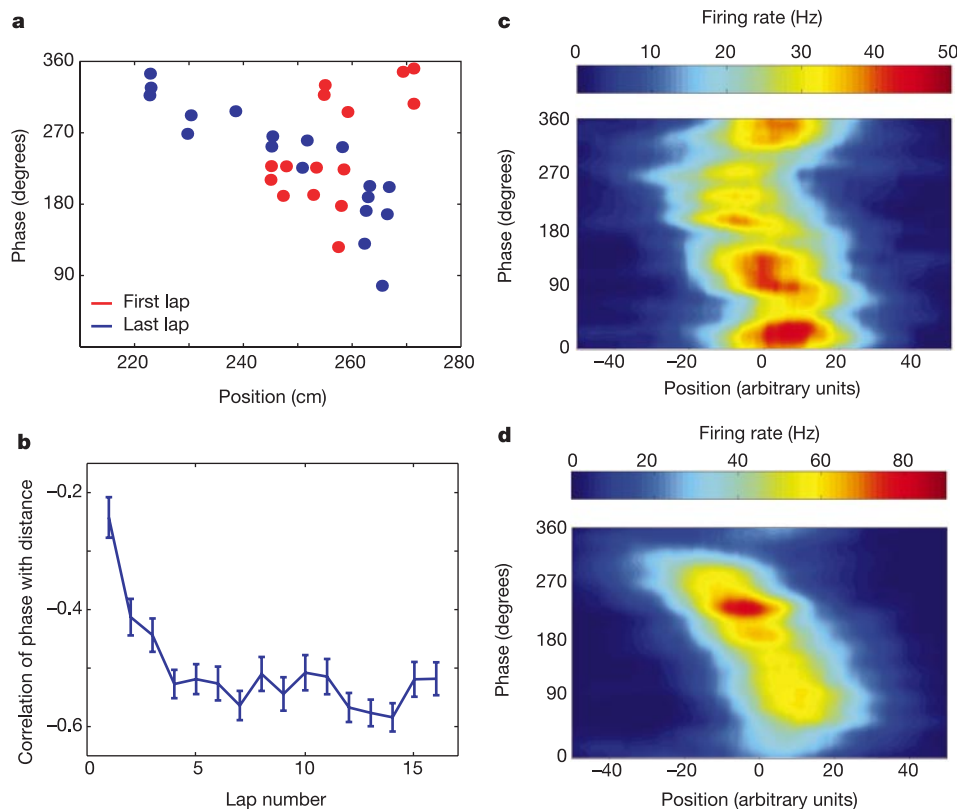


Figure 4 Experience dependence of the temporal code. **a**, All the spikes fired by a neuron (same as in Fig. 1a) during the first lap (red dots) through the place field and during the last (thirty-first) lap (blue dots). Although the phase is positively correlated with position in the first lap ($r = 0.6$), it is more strongly and negatively correlated with position after experience ($r = -0.9$). **b**, Correlation of phase with position was calculated for each cell for each lap in which the cell fired more than two spikes. The mean value of the correlation of phase with position, averaged across the population of place cells, is plotted as a function of experience. The phase is 2.1 times more strongly correlated with position in

the sixteenth lap (mean $r = -0.50$) compared to the first lap (mean $r = -0.24$, $P < 0.0001$). **c**, **d**, Population-averaged STRF (see Methods) is plotted for the first lap (**c**) and the sixteenth lap (**d**). The mean phase is 2 times more strongly correlated with position in the sixteenth lap than in the first lap, and the mean phase is larger. The mean value of the phase of all the spikes in a given lap increased with experience from 171° before experience to 185° after experience, and the distribution of phases became wider with experience.

is turned on, the latency to first spike fired by a neuron would be inversely proportional to the subsequent firing rate of the neuron. Thus the latency of the first spike could provide rapid information about the stimulus. In the presence of oscillations this information would be repeated in each cycle.

One of the critical tasks of the central nervous system is to learn the causal relationships between events^{12,15,26,27}. For example, spatial navigation may require formation of memory for the temporal order of activation of spatially selective cells^{12,15,27–29}. But whereas the receptive fields are activated in a given sequence on a behaviourally relevant timescale ($>1,000$ ms), the biophysical laws of plasticity^{1,2} require that the same sequence of events be replayed on short (<10 ms) timescales. This will not occur consistently if the neurons fired purely in a rate-varying Poisson fashion. However, in the presence of oscillations and asymmetric excitation, two place cells that are sequentially activated several hundred milliseconds apart will be activated in the same temporal sequence within tens of milliseconds, resulting in a replay (that is, binding) of the sequence within one theta cycle, and hence rapid learning of temporal sequences.

Thus oscillations can transform an asymmetric rate code into a temporal code. This can have a critical role in temporal sequence learning by compressing and replaying the behaviourally relevant temporal order of events occurring on long, physiological timescales into short timescales^{5,6,27} relevant for synaptic plasticity. □

Methods

Three Long-Evans rats were trained to run on linear tracks (see refs 12 and 29 for details), and single unit data were recorded from the CA1 region of the dorsal hippocampus using tetrodes¹² according to NIH guidelines. Local field potentials (sampled at 2 kHz, filtered between 0.1 Hz and 900 Hz) and spike data (sampled at 31 kHz and filtered between 300 Hz to 9 kHz) were recorded from the same tetrodes in the pyramidal layer, along with the rat's position (spatial resolution 0.66 cm, sampling rate 30 Hz) and head direction. A total of 238 cells were active during behaviour, and were recorded in 3 rats in 7 sessions. Data from 158 of these cells were used in a previous study¹². The tracks were 'linearized' for the purpose of analysis such that the distance increased along the rat's direction of motion through the place field. Data from goal locations were not used^{12,29}. 171 place fields obtained from 140 place cells were used for the analyses.

The local field potential was bandpass-filtered off-line in the theta band (between 4 and 14 Hz), and the peaks and troughs of the theta rhythm were detected. Spike phase was computed with respect to the troughs of the filtered local field potential with the highest theta modulation (Fig. 2), with a phase offset that provided the highest correlation between phase and position⁵. A large proportion (82%) of neurons exhibited strongest phase precession with a negative phase offset (population average, -30°). All the subsequent analyses were carried out with respect to this ideal phase origin for each neuron.

Calculations involving mean and standard deviation of phases were done using circular statistics³⁰. Thus, the value of width of a phase distribution³⁰ ranged between 0 and $\sqrt{2}$. Correlation of phase and position was computed using linear statistics. Mean values were computed over the entire population of 171 place fields, and their significance was estimated with Student's *t*-test. In order to compute the relationship between the rate and phase, spikes were arranged in ascending order of location. Mean position, mean rate, mean phase and the width of phase distribution, normalized by occupancy, were calculated for successive blocks of 5% of spikes. The last bin therefore had a variable number of spikes (between 0 and 5%) and hence was not used for the subsequent analyses. Relationships between these variables were obtained by computing the correlation coefficient across the blocks of data.

Hippocampal spatio-temporal receptive fields were computed by dividing the number

of spikes fired by a neuron in each spatio-temporal bin (spatial width 1 pixel = 0.66 cm, temporal width 1° ≈ 0.35 ms) by the total time spent by the rat in that bin. The result was smoothed by convolution with a two-dimensional gaussian (spatial width 2.4 cm, temporal width 12° ≈ 4.2 ms).

Received 24 October 2001; accepted 22 April 2002; doi:10.1038/nature00807.

- Markram, H., Lübke, J., Frotscher, M. & Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**, 213–215 (1997).
- Bi, G. Q. & Poo, M. M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
- O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
- Wilson, M. A. & McNaughton, B. L. Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993).
- O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
- Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
- Hopfield, J. J. Pattern recognition computation using action potential timing for stimulus representation. *Nature* **376**, 33–36 (1995).
- Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J. & McNaughton, B. L. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* **6**, 271–280 (1996).
- Wallenstein, G. V. & Hasselmo, M. E. GABAergic modulation of hippocampal population activity: sequence learning, place field development, and the phase precession effect. *J. Neurophysiol.* **78**, 393–408 (1997).
- Kamondi, A., Acsády, L., Wang, X. J. & Buzsáki, G. Theta oscillations in somata and dendrites of hippocampal pyramidal cells in vivo: activity-dependent phase-precession of action potentials. *Hippocampus* **8**, 244–261 (1998).
- Bose, A., Booth, V. & Recce, M. A temporal mechanism for generating the phase precession of hippocampal place cells. *J. Comput. Neurosci.* **9**, 5–30 (2000).
- Mehta, M. R., Quirk, M. C. & Wilson, M. A. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* **25**, 707–715 (2000).
- Toth, K., Freund, T. F. & Miles, R. Disinhibition of rat hippocampal pyramidal cells by GABAergic afferents from the septum. *J. Physiol.* **500**, 463–474 (1997).
- Mehta, M. R. & Wilson, M. A. From hippocampus to V1: Effect of LTP on spatio-temporal dynamics of receptive fields. *Neurocomputing* **32**, 905–911 (2000).
- Mehta, M. R. Neuronal dynamics of predictive coding. *Neuroscientist* **7**, 490–495 (2001).
- Buzsáki, G., Rappelsberger, P. & Kelenyi, L. Depth profiles of hippocampal rhythmic slow activity ('theta rhythm') depend on behaviour. *Electroencephalogr. Clin. Neurophysiol.* **61**, 77–88 (1985).
- Jensen, O. & Lisman, J. E. Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells. *Learn. Mem.* **3**, 279–287 (1996).
- Hasselmo, M. E., Fransen, E., Dickson, C. & Alonso, A. A. Computational modeling of entorhinal cortex. *Ann. NY Acad. Sci.* **911**, 418–446 (2000).
- Hirase, H., Czurko, H. H., Csicsvari, J. & Buzsáki, G. Firing rate and theta-phase coding by hippocampal pyramidal neurons during 'space clamping'. *Eur. J. Neurosci.* **11**, 4373–4380 (1999).
- Ekstrom, A. D., Meltzer, J., McNaughton, B. L. & Barnes, C. A. NMDA receptor antagonism blocks experience-dependent expansion of hippocampal "place fields". *Neuron* **31**, 631–638 (2001).
- Magee, J. C. Dendritic mechanisms of phase precession in hippocampal ca1 pyramidal neurons. *J. Neurophysiol.* **86**, 528–532 (2001).
- Livingstone, M. S. Mechanisms of direction selectivity in macaque V1. *Neuron* **20**, 509–526 (1998).
- Chance, F. S., Nelson, S. B. & Abbott, L. F. Synaptic depression and the temporal response characteristics of V1 cells. *J. Neurosci.* **18**, 4785–4799 (1998).
- Rao, R. P. & Sejnowski, T. J. Predictive learning of temporal sequences in recurrent neocortical circuits. *Novartis Found. Symp.* **239**, 208–229 (2001).
- König, P., Engel, A. K., Roelfsema, P. R. & Singer, W. How precise is neuronal synchronization? *Neural Comput.* **7**, 469–485 (1995).
- Levy, W. B., in *Computational Models of Learning in Simple Neural Systems* (eds Hawkins, R. D. & Bower, J. H.) 243–305 (Academic, New York, 1989).
- Blum, K. I. & Abbott, L. F. A model of spatial map formation in the hippocampus of the rat. *Neural Comput.* **8**, 85–93 (1996).
- Mehta, M. R. & McNaughton, B. L. in *Computational Neuroscience: Trends in Research* (ed. Bower, J.) 741–745 (Plenum, New York, 1996).
- Mehta, M. R., Barnes, C. A. & McNaughton, B. L. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc. Natl Acad. Sci. USA* **94**, 8918–8921 (1997).
- Zar, J. H. *Biostatistical Analysis* (Prentice Hall, New Jersey, 1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank W.F. Asaad, G. Liu, K. Louie, J. Raymond and S. Schnall for comments on the manuscript. This work was supported by the NIH (M.A.W.) and an HHMI pre-doctoral fellowship (A.K.L.). Parts of this work were presented at the Computational Neuroscience Meeting 2000 and at the Society for Neuroscience meeting 2001.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.R.M. (e-mail: mayank@mit.edu) or M.A.W. (e-mail: wilson@cortical.mit.edu).

Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*

Nathan Shankar*, Arto S. Baghdayan* & Michael S. Gilmore†

* Department of Pharmaceutical Sciences and † Departments of Ophthalmology and Microbiology & Immunology, University of Oklahoma Health Sciences Center, PO Box 26901, Oklahoma City, Oklahoma 73190, USA

Enterococci are members of the healthy human intestinal flora, but are also leading causes of highly antibiotic-resistant, hospital-acquired infection¹. We examined the genomes of a strain of *Enterococcus faecalis* that caused an infectious outbreak in a hospital ward in the mid-1980s (ref. 2), and a strain that was identified as the first vancomycin-resistant isolate in the United States³, and found that virulence determinants were clustered on a large pathogenicity island, a genetic element previously unknown in this genus. The pathogenicity island, which varies only subtly between strains, is approximately 150 kilobases in size, has a lower G + C content than the rest of the genome, and is flanked by terminal repeats. Here we show that subtle variations within the structure of the pathogenicity island enable strains harbouring the element to modulate virulence, and that these variations occur at high frequency. Moreover, the enterococcal pathogenicity island, in addition to coding for most known auxiliary traits that enhance virulence of the organism, includes a number of additional, previously unstudied genes that are rare in non-infection-derived isolates, identifying a class of new targets associated with disease which are not essential for the commensal behaviour of the organism.

In a study of a hospital ward epidemic in the mid-1980s, one *E. faecalis* strain resistant to many different antibiotics was identified as having caused over 30 infections (typified by isolate MMH594) and carried a fivefold increased risk of death within 3 weeks of isolation². Several years later, the first vancomycin-resistant enterococcal isolate in the United States was identified approximately 400 miles away, and serial isolates (V583 and V586) were obtained from a chronically infected patient³. The vancomycin-resistance determinant has been characterized in molecular detail⁴, and strain V583 was provided for genome analysis (for strain V583 genome data see <http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=gef>).

We and others have analysed a number of virulence factors in *E. faecalis*, including a structurally novel toxin, the cytolysin^{5–7}, and a surface protein, Esp⁸, which contributes to colonization of the bladder in a model of urinary tract infection⁹. Esp confers biofilm production capability to enterococci¹⁰, and a highly conserved Esp variant is common among clinical vancomycin-resistant and vancomycin-sensitive *E. faecium* strains^{11–13}. *Enterococcus faecalis* strains V583 and V586 were isolated from a patient 11 days apart³, and initial studies failed to identify differences between these isolates as they exhibited what appeared to be identical DNA fingerprint patterns. However, subsequent experiments designed to localize genes coding for both Esp and cytolysin found them on the chromosome of V586, but not V583. The difference was localized to a relatively small deletion within a roughly 315 kilobase (kb) SfiI fragment in V583 (Fig. 1), which suggested that *cyl* and *esp* are closely linked.

Alignment and comparison of nucleotide sequences coding for the cytolysin operon, *esp* and surrounding regions of the V586 genome (Esp⁺, Cyl⁺), and the point of deletion of these functions in the genome of V583 (Esp[–], Cyl[–]), (see <http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=gef>), revealed a

17,036-bp deletion in V583 (Fig. 2a), extending from within EF0047–EF0057 (see Supplementary Information Table 1). In the genome of V586, IS256- and IS905-like insertion elements (refs 14 and 15, respectively) occur within the cytolysin operon⁶ in *cylB*, resulting in strain V586 being phenotypically non-cytolytic, although it harbours the entire cytolysin operon. An additional IS905-like element occurs in strain V586 at the distal end of the 17,036 bp segment, bracketing the *esp* gene and several other functions of unknown relevance to disease (Fig. 2a). Strain V583 (*Esp*[−], *Cyl*[−]) was generated by spontaneous deletion of DNA from the V586 genome, beginning at the point of fusion between the IS256-like element and *cylB*, and ending at a point 17,036-bp downstream that occurs 276 bp in the 3' direction to the distal IS905-like element, between direct repeats of a 5-bp sequence (Fig. 2b). The deletion resulting in strain V583 removed the 3' half of the cytolysin operon and regions downstream, including the entire *esp* gene, rendering this strain deficient in expression of either of these virulence traits.

Quantitative polymerase chain reaction (PCR) revealed that precise excision of the 17,036-bp DNA segment from strain V586 (*Esp*⁺, *Cyl*[−]) occurs at a frequency of approximately 1 in 10³ cells during overnight cultivation *in vitro*, a rate similar to that observed in modulation of virulence in other bacterial species by phase variation¹⁶. The mechanism for this specific, high frequency recombination is presently unknown, but it is suggestive of a highly evolved switch for modulating bacterial virulence. Strain V583 (*Esp*[−], *Cyl*[−]) was the first of serial isolates, and is the strain that has undergone deletion. This suggests that either the patient was infected with a population of *E. faecalis* that was heterogeneous with respect to this deletion, or that deletion occurred on subsequent laboratory passage.

As *esp* and the cytolysin operon were co-localized on the chromosome of the isolate that had not undergone this spontaneous deletion, strain V586, it was of interest to determine whether they were also linked on the genome of strain MMH594

(*Esp*⁺, *Cyl*⁺), which caused multiple life-threatening infections at another healthcare centre several years earlier². Hybridization and single-pass sequencing revealed that the region spanning cytolysin and *esp* genes on strain V586 (Fig. 2a) was nearly identical on the genome of MMH594, except for the absence of the IS256- and IS905- like elements that interrupt *cylB* in strain V586. MMH594 is therefore phenotypically *Esp*-positive and cytolytic, and seems to represent the prototype from which strains V586 and V583 evolved.

Nucleotide sequence similarities between the genomes of strain V586 and MMH594 did not end in the immediate vicinity of the cytolysin and *esp* genes, however, suggesting the presence of a larger common element on which these virulence traits reside. Primer walking was used to extend MMH594 and V586 sequence analysis in both directions outward from the point where *Esp* and cytolysin determinants were deleted in V583, until sequences that were shared by a non-infection-derived oral isolate of *E. faecalis*, strain OG1 (ref. 17), were encountered. These experiments identified an insertion of 153,571 bp of DNA in the genome of the clinical isolate MMH594 (GenBank accession number AF454824) with highly conserved variations in V586 and V583, which was absent from strain OG1. The point at which this large genetic element had inserted into the chromosomes of the clinical isolates occurred 3' to a 10-bp sequence on the genome that was reiterated at the distal end of the element (Fig. 3a). Comparison of the element in MMH594 with those of V586 and V583 (Fig. 3b) revealed the accretion of additional IS elements in the latter isolates, and a 2.8-kb region

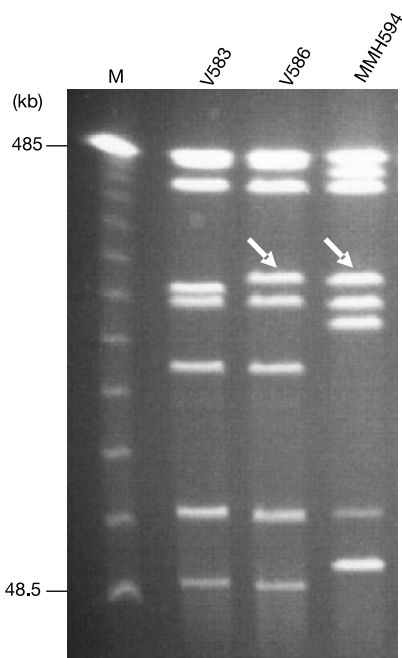


Figure 1 Pulsed-field gel electrophoretic (PFGE) banding patterns of *Sfi*I-restricted genomic DNA⁸. Arrows point to the approximately 315-kb restriction fragment from V586 and MMH594 DNA to which probes for *esp* and *cylA* hybridize. Lane M is Lambda Ladder PFG marker from New England Biolabs.

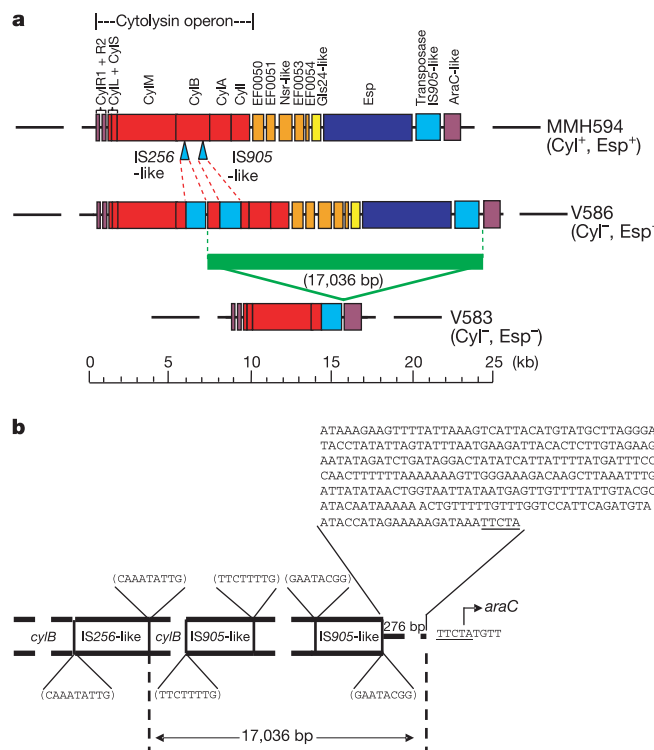


Figure 2 Schematic alignment and comparison of the genomic regions harbouring *Esp*, cytolysin and flanking functions within the PAI in MMH594, to similar regions in strains V586 and V583. **a**, The insertion sites for IS256- and IS905-like elements on the MMH594 chromosome, (insertions which result in the V586 phenotype) are shown by solid triangles. The region bounded by the dotted vertical lines denotes the 17,036-bp DNA segment that spontaneously excises out of the V586 chromosome resulting in V583. **b**, Junction sequences on the chromosome of V586 at the site of deletion of the 17,036-bp DNA segment. The 9- and 8-bp target sites duplicated after the insertion of the IS256- and IS905-like sequences are shown in parentheses. The 5-bp direct repeat at the right junction is underlined.

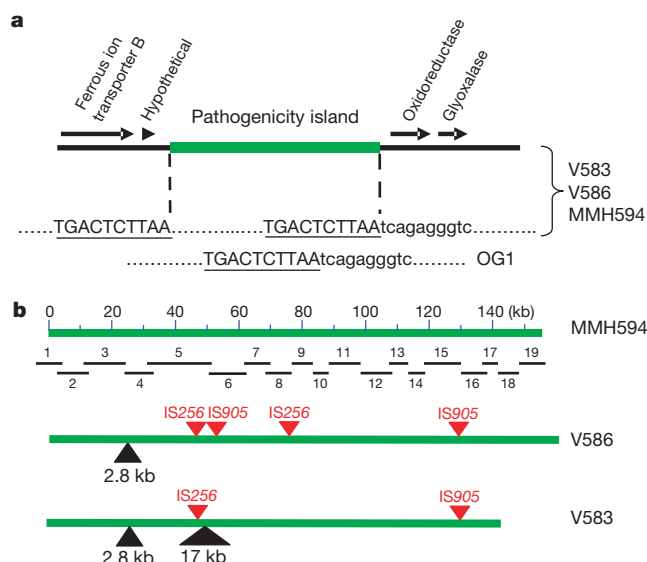


Figure 3 Junction sequences flanking the PAI in infection-derived isolates V583 (Esp⁻, Cyl⁻), V586 (Esp⁺, Cyl⁻) and MMH594 (Esp⁺, Cyl⁺). **a**, Primer pair PAI164 and PAI167 (see Methods) was used to PCR amplify a 2,121-bp segment of OG1 DNA spanning the predicted PAI insertion site and the product sequenced to confirm the point of insertion. The 10-bp duplicated sequence is underlined. **b**, Schematic alignment and comparison of

the PAI in the three clinical isolates. The red triangles denote the locations of IS elements in strains V583 and V586, which are absent in MMH594. The black triangles represent regions present in MMH594 but missing from the V583 and V586 genomes. The 2.8-kb region specifies the location of a group II intron-like sequence. Numbered horizontal bars indicate overlapping regions of the PAI amplified from genomic DNA.

unique to MMH594. Notably, the 2.8-kb region in MMH594 possesses the features of a group II intron including consensus 5' (GTGCG) and 3' (AT) splice sites¹⁸, and a single reading frame coding for a homologue of the intron-coded protein IepA¹⁹. Excision of this intron would be predicted to generate an in-frame fusion of the 5' and 3' exons of *traG*²⁰ in strains V583 and V586. In contrast to the high-frequency deletion of the Esp and cytolysin-coding region from within the pathogenicity island (PAI) of V586, we determined by quantitative PCR that the PAI itself is integrated in a stable manner in the chromosomes of V583, V586 and MMH594, and does not undergo spontaneous deletion within the limits of detection.

The genetic element in MMH594 codes for 129 open reading frames (ORFs) (Fig. 4) and possesses the hallmarks of a PAI²¹ including large size, terminal duplication at the target site, significantly variant G + C content of 32.2%, and the presence of genes that code for transposases, transcriptional regulators and proteins with known or potential roles in virulence or adaptation and survival in different environments. Virulence traits coded for within the *E. faecalis* PAI (Fig. 4 and Supplementary Information Table 1) include the aforementioned Esp and cytolysin, and also aggregation substance²², which contributes to bacterial clumping, survival in neutrophils²³ and adherence to host tissues²⁴. Additionally, within the element are genes that seem to code for a DNA-damage-inducible protein, an AraC-like transcriptional regulator, a conjugated bile acid hydrolase, components of the phosphotransferase system, and a Gls24-like starvation-inducible protein²⁵; however, there are none that are known to code for antibiotic resistance. The PAI also codes for 18 ORFs for which no function could be predicted. The role of these genes and bacterial survival or proliferation in the unique environment encountered in a hospital, or at sites of infection, remains to be explored.

To test independently whether new traits not previously associated with known roles in enterococcal disease pathogenesis were truly elements of a pathogenicity island, or were randomly interspersed chromosomal genes of *E. faecalis*, we performed the

following experiment. A panel of 40 clinical isolates from widely dispersed geographical sites, and 40 fecal isolates from healthy volunteers, were examined by hybridization for the presence of the detected *araC*-like regulator, the stress-inducible *gls24* homologue, and the inferred conjugated bile acid hydrolase (*cbh*). Clinical isolates were observed to be significantly enriched for all three genes: *araC*-like ($P < 0.001$), *cbh* ($P < 0.01$) and *gls24*-like ($P < 0.001$).

Unlike a large number of bacterial PAIs that insert within transfer RNA genes²¹, the *E. faecalis* PAI occurs in a non-coding region flanked by an ORF specifying a hypothetical protein with no known homologue or predictable function on one side, and a putative oxidoreductase of unknown function but related to the aldo/keto reductase family, on the other (Fig. 3a). ORFs in the vicinity of the insertion site include those specifying putative ferrous ion transport functions and a glyoxalase family protein, but none that would suggest a known insertion selection mechanism.

Extensive nucleotide sequence identity exists at the 5' end of the PAI with contiguous conjugation-related structural genes of enterococcal pheromone-responsive plasmids pAM373 and pAD1 (ref. 20). This suggests that at least one-third of the PAI evolved from the integration of such plasmid sequences into the chromosome, although the nucleotide sequence flanking the apparent plasmid-derived regions failed to provide a discernable mechanism for integration. Subsequent divergence apparently arose from the insertion of transposable elements that distance the cytolysin determinants from the conjugation-related structural genes. The only transfer-related genes present, however, are those specifying a TraG-like protein (unknown function) and a region with 87% identity at the nucleotide level to a second transfer origin (*oriT*) recently identified in pAD1 (ref. 20), suggesting the prospect for mobilization of this element from a donor to recipient strain. Filter mating experiments revealed that although the erythromycin-resistant methylase (Erm^r) and gentamicin resistance (Gm^r) determinants could be transferred from MMH594b (donor, derivative of MMH594 with the PAI tagged with a chloramphenicol resistance (Cm^r) determinant) to FA2-2 (plasmid-free recipient, lacking the

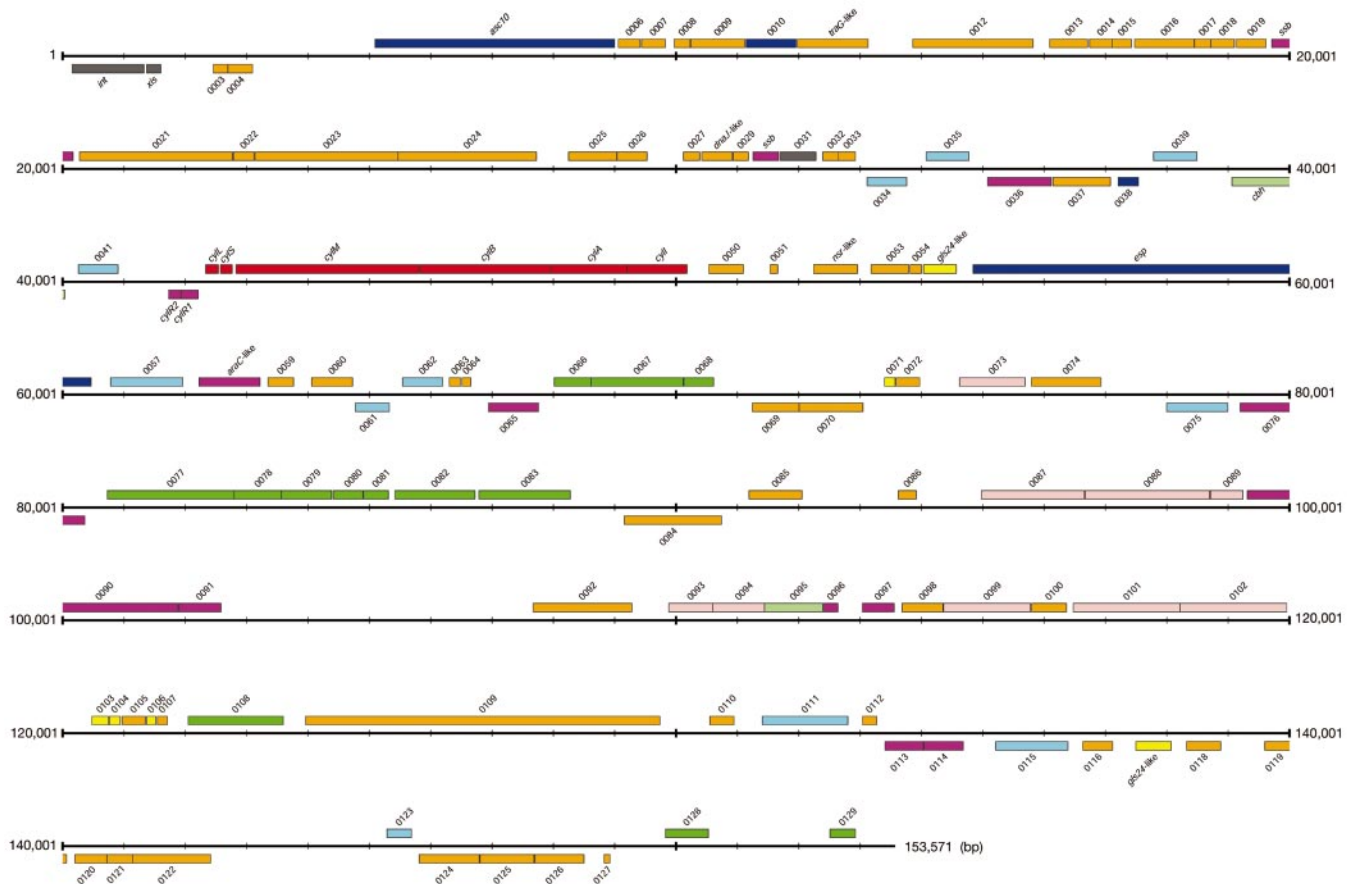


Figure 4 Linear representation of the pathogenicity island in *E. faecalis* MMH594. Coding sequences were defined by combining predictions made by Artemis²⁹ with visual inspection of ORFs for the presence of start codons and upstream ribosome-binding sites. Potential coding sequences were then subjected to BLASTp similarity searches against the non-redundant protein database. Predicted coding regions transcribed on the forward DNA strand are displayed above the genomic scale line and those transcribed on the reverse DNA strand are displayed below the scale line. Proposed gene name designations

or gene identification numbers are presented with each coding region. Colour coding of the ORFs is as follows: dark grey, integrase, excisionase, endonuclease; red, cytotoxin operon; dark blue, cell surface associated; orange, hypothetical, conserved hypothetical; light green, lipid metabolism, lipoprotein, bile acid hydrolase; magenta, transcriptional regulator, repressor, single-strand binding protein, signal transduction; light blue, transposase/IS elements, recombinase; dark green, carbohydrate metabolism, modifier; yellow, ribosomal, stress-response; light pink, ABC transporter, ATP-binding.

pathogenicity island) at frequencies of 5.02×10^{-6} and 2.6×10^{-8} per donor cell, respectively, there was no evidence for transfer of the *Cm^r* marker into FA2-2, suggesting that transfer of the PAI occurs more on an evolutionary scale rather than on a laboratory scale.

Although chromosomal elements bearing some of the hallmarks of PAIs have been identified in *Staphylococcus aureus*²⁶ and *Streptococcus pneumoniae*²⁷, these lack several of the features common in PAIs, such as large size, variant G + C content, or known relationship to virulence. The 150-kb PAI identified within multiple antibiotic-resistant pathogenic isolates of *E. faecalis* seems to be the first element from a Gram-positive organism to satisfy all criteria.

At least one of the traits found within this pathogenicity island, Esp, is also highly associated with infection-derived isolates of *E. faecium*. The discovery of a number of ORFs with no assignable function included within the PAI, which are absent from a non-infection-derived isolate and rare in fecal isolates, implicates their contribution to enterococcal survival in the hospital, or in the process of disease transmission or pathogenesis. This PAI, or elements thereof, may be a useful marker for detecting unusually pathogenic enterococcal strains, triggering enhanced infection control procedures within a hospital ward. In addition to known virulence traits coded for by this PAI, the PAI circumscribes a

cluster of new genes around which therapeutic interventions may be designed. □

Methods

DNA manipulations

Preparation of enterococcal DNA, analysis by restriction mapping, inverse-PCR, chromosome walking, Southern hybridization, nucleotide sequence determination and alignment was essentially done following previously published⁸ and standard protocols²⁸.

Nucleotide sequence information from strains MMH594 and V586, throughout regions conserved in V583, was obtained by single-pass sequencing and verified by restriction analysis. Custom oligonucleotide primers (Integrated DNA Technologies) based on the emerging nucleotide sequence of strain V583 was used for PCR amplification as well as sequencing of gel-purified PCR amplification products. Takara LA PCR kit (Panvera) was used for all PCR reactions. Nucleotide sequence information from regions unique to strains MMH594 and V586 was obtained by direct sequencing of both strands of PCR products using a primer walking approach. Where necessary, restriction fragments from PCR products were cloned into pBluescript II SK⁻ (Stratagene) and sequenced.

For alignment and comparison of the PAI sequences in the three clinical isolates, overlapping regions (segments 1–19 (Fig. 3b)) of the PAI were amplified from genomic DNA, the amplification product in each case was gel purified, restriction mapped using enzymes *EcoRI* and *HindIII*, and sequenced using a primer walking approach. A total of 219 oligonucleotide primers were used to obtain the nucleotide sequence of the PAI in MMH594 and V586.

Quantitative PCR

For quantitative PCR to estimate the frequency of deletion of the 17,036-bp region from within the PAI in V586, genomic DNA prepared from 9×10^8 colony-forming units

(c.f.u.) of V586 was resuspended in 200 µl of TE buffer. Serial dilutions of 1:1, 1:2.5, 1:5, 1:10 and 1:20 were made in TE buffer and 5 µl aliquots representing the DNA concentrations 875 ng, 700 ng, 350 ng, 175 ng and 87.5 ng were subjected to 40 cycles of PCR amplification with the primer pair L1 (5'-GGAGGACTAACTATCAAGAATATT CGTAC-3') and R1 (5'-CAGAACGTACGTTATCCCCACTAACACACAGGA-3'). This primer pair amplifies a 2.4-kb region spanning the deletion site only in strain V583. The lowest concentration of V583 template that resulted in a detectable product under the PCR conditions used was 175 pg of template (equivalent to DNA from approx. 2.2×10^7 c.f.u.). A detectable amplification product of similar intensity was obtained from 175 ng of V586 DNA (equivalent to DNA from approx. 2.2×10^6 c.f.u.), demonstrating that the frequency of deletion of the 17,036-bp DNA segment from V586 is approximately 1 in 10^3 cells.

Quantitative PCR to determine the frequency of deletion of the entire PAI was done essentially as described above using primer pair PAI164 (5'-ATGCCATGTTTCAGC GAAGTTGCCAATTATC-3') and PAI167 (5'-GCTGATTATATGGTTCTCAGC AATCGCC-3'). The lowest concentration of OG1 DNA (which lacks the PAI) that resulted in an amplified product of 2,121 bp was 20 pg (equivalent to DNA from 2×10^2 c.f.u.). No amplified product was detected when up to 2 µg (equivalent to DNA from 2×10^7 c.f.u.) of genomic DNA from V583, V586 or MMH594 was used as template, indicating that deletion of the entire PAI in these strains occurs at a frequency of less than 1 in 10^5 cells.

Received 31 December 2001; accepted 3 April 2002; doi:10.1038/nature00802.

1. Richards, M. J., Edwards, J. R., Culver, D. H. & Gaynes, R. P. Nosocomial infections in combined medical-surgical intensive care units in the United States. *Infect. Control Hosp. Epidemiol.* **21**, 510–515 (2000).

2. Huycke, M. M., Spiegel, C. A. & Gilmore, M. S. Bacteremia caused by hemolytic, high-level gentamicin-resistant *Enterococcus faecalis*. *Antimicrob. Agents Chemother.* **35**, 1626–1634 (1991).

3. Sahm, D. F. *et al.* *In vitro* susceptibility studies of vancomycin-resistant *Enterococcus faecalis*. *Antimicrob. Agents Chemother.* **33**, 1588–1591 (1989).

4. Arthur, M. & Courvalin, P. Genetics and mechanisms of glycopeptide resistance in enterococci. *Antimicrob. Agents Chemother.* **37**, 1563–1571 (1993).

5. Jett, B. D., Huycke, M. M. & Gilmore, M. S. Virulence of enterococci. *Clin. Microbiol. Rev.* **7**, 462–478 (1994).

6. Haas, W., Shepard, B. D. & Gilmore, M. S. Two-component regulator of *Enterococcus faecalis* cytolysin responds to quorum-sensing autoinduction. *Nature* **415**, 84–87 (2002).

7. Ike, Y., Hashimoto, H. & Clewell, D. B. Hemolysin of *Streptococcus faecalis* subspecies *zymogenes* contributes to virulence in mice. *Infect. Immun.* **45**, 528–530 (1984).

8. Shankar, V., Baghdadyan, A. S., Huycke, M. M., Lindahl, G. & Gilmore, M. S. Infection-derived *Enterococcus faecalis* strains are enriched in *esp*, a gene encoding a novel surface protein. *Infect. Immun.* **67**, 193–200 (1999).

9. Shankar, N. *et al.* Role of *Enterococcus faecalis* surface protein *Esp* in the pathogenesis of ascending urinary tract infection. *Infect. Immun.* **69**, 4366–4372 (2001).

10. Toledo-Arana, A. *et al.* The enterococcal surface protein, *Esp*, is involved in *Enterococcus faecalis* biofilm formation. *Appl. Environ. Microbiol.* **67**, 4538–4545 (2001).

11. Willems, R. J. L. *et al.* Variant *esp* gene as a marker of a distinct genetic lineage of vancomycin-resistant *Enterococcus faecium* spreading in hospitals. *Lancet* **357**, 853–855 (2001).

12. Woodford, N., Soltani, M. & Hardy, K. J. Frequency of *esp* in *Enterococcus faecium* isolates. *Lancet* **358**, 584 (2001).

13. Baldassarri, L., Bertuccini, L., Ammendolia, M. G., Gherardi, G. & Creti, R. Variant *esp* gene in vancomycin-sensitive *Enterococcus faecium*. *Lancet* **357**, 1802 (2001).

14. Rouch, D. A., Byrne, M. E., Kong, Y. C. & Skurray, R. A. The *aacA-aphD* gentamicin and kanamycin resistance determinant of Tn4001 from *Staphylococcus aureus*: expression and nucleotide sequence analysis. *J. Gen. Microbiol.* **133**, 3039–3052 (1987).

15. Dodd, H. M., Horn, N. & Gasson, M. J. Characterization of IS905, a new multicopy insertion sequence identified in lactococci. *J. Bacteriol.* **176**, 3393–3396 (1994).

16. Fierer, J. & Guiney, D. G. Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. *J. Clin. Invest.* **107**, 775–780 (2001).

17. Gold, O. G., Jordan, H. V. & van Houtte, J. The prevalence of enterococci in the human mouth and their pathogenicity in animal models. *Arch. Oral Biol.* **20**, 473–477 (1975).

18. Ferat, J. L. & Michel, F. Group II self-splicing introns in bacteria. *Nature* **364**, 358–361 (1993).

19. Huang, C. C., Narita, M., Yamagata, T. & Endo, G. Identification of three *merB* genes and characterization of a broad-spectrum mercury resistance module encoded by a class II transposon of *Bacillus megaterium* strain MB1. *Gene* **239**, 361–366 (1999).

20. Francia, M. V. *et al.* Completion of the nucleotide sequence of the *Enterococcus faecalis* conjugative virulence plasmid pAD1 and identification of a second transfer origin. *Plasmid* **46**, 117–127 (2001).

21. Hacker, J. & Kaper, J. B. Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**, 641–679 (2000).

22. Galli, D. & Wirth, R. Comparative analysis of *Enterococcus faecalis* sex pheromone plasmids identifies a single homologous DNA region which codes for aggregation substance. *J. Bacteriol.* **173**, 3029–3033 (1991).

23. Rakita, R. M. *et al.* *Enterococcus faecalis* bearing aggregation substance is resistant to killing by human neutrophils despite phagocytosis and neutrophil activation. *Infect. Immun.* **67**, 6067–6075 (1999).

24. Rozdzinski, E., Marre, R., Susa, M., Wirth, R. & Muscholl-Silberhorn, A. Aggregation substance-mediated adherence of *Enterococcus faecalis* to immobilized extracellular matrix proteins. *Microb. Pathog.* **30**, 211–220 (2001).

25. Giard, J. C., Rince, A., Capioux, H., Auffray, Y. & Hartke, A. Inactivation of the stress- and starvation-inducible *gls24* operon has a pleiotropic effect on cell morphology, stress sensitivity, and gene expression in *Enterococcus faecalis*. *J. Bacteriol.* **182**, 4512–4520 (2000).

26. Novick, R. P., Schlievert, P. & Ruzin, A. Pathogenicity and resistance islands of staphylococci. *Microbes Infect.* **3**, 585–594 (2001).

27. Brown, J. S., Gilliland, S. M. & Holden, D. W. A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. *Mol. Microbiol.* **40**, 572–585 (2001).

28. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1990).

29. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

This work was supported by grants from the National Institutes of Health, American Heart Association and Research to Prevent Blindness. We thank M. Carson for help with generating the linear map of the pathogenicity island.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.S. (e-mail: nathan-shankar@ouhsc.edu).

Subendothelial retention of atherogenic lipoproteins in early atherosclerosis

Kristina Skålen*†, Maria Gustafsson*†, Ellen Knutsen Rydberg*, Lillemor Mattsson Hultén*, Olov Wiklund*, Thomas L. Innerarity‡ & Jan Borén*

* Wallenberg Laboratory for Cardiovascular Research, Göteborg University, Göteborg S-41345, Sweden

‡ Gladstone Institute of Cardiovascular Disease, San Francisco, California 94141-9100, USA

† These authors contributed equally to this work

Complications of atherosclerosis are the most common cause of death in Western societies¹. Among the many risk factors identified by epidemiological studies, only elevated levels of lipoproteins containing apolipoprotein (apo) B can drive the development of atherosclerosis in humans and experimental animals even in the absence of other risk factors². However, the mechanisms that lead to atherosclerosis are still poorly understood. We tested the hypothesis that the subendothelial retention of atherogenic apoB-containing lipoproteins is the initiating event in atherogenesis³. The extracellular matrix of the subendothelium, particularly proteoglycans, is thought to play a major role in the retention of atherogenic lipoproteins⁴. The interaction between atherogenic lipoproteins and proteoglycans involves an ionic interaction between basic amino acids in apoB100 and negatively charged sulphate groups on the proteoglycans⁵. Here we present direct experimental evidence that the atherogenicity of apoB-containing low-density lipoproteins (LDL) is linked to their affinity for artery wall proteoglycans. Mice expressing proteoglycan-binding-defective LDL developed significantly less atherosclerosis than mice expressing wild-type control LDL. We conclude that subendothelial retention of apoB100-containing lipoprotein is an early step in atherogenesis.

Table 1 Mutants of the human apoB100 gene

Recombinant LDL	LDL receptor binding	Proteoglycan binding
Control	Normal	Normal
W ₄₃₆₉ → Y	Defective	Normal
R _{K3359-3369} → S _A	Defective	Defective
K ₃₃₆₃ → E	Normal	Defective
6-GBSM	Defective	Defective

To investigate the potential importance of LDL binding to artery wall proteoglycans in atherogenesis, we created transgenic mice expressing five types of human recombinant LDL (Table 1), fed them an atherogenic diet (1.2% cholesterol, 0.5% cholic acids, and 18% fat) for 20 weeks, and quantitated the extent of atherosclerosis. The first transgenic mouse line expressed human recombinant control LDL⁶. The second transgenic mouse line expressed recombinant LDL with a tyrosine substituted for tryptophan-4369 in apoB100 ($W_{4369} \rightarrow Y$), causing a conformational change that disrupts LDL-receptor binding but does not affect the interaction with proteoglycans⁵. The ability to discriminate between proteoglycan-binding activity and LDL-receptor-binding activity was important because the principal proteoglycan-binding site coincides with the LDL-receptor-binding site in apoB100⁶.

The third transgenic mouse line expressed recombinant LDL in which the basic amino acids in the proteoglycan-binding site of apoB (residues 3359–3369) were converted to neutral amino acids: arginines to serines and lysines to alanines ($R,K_{3359-3369} \rightarrow S,A$)⁶. The fourth expressed human recombinant LDL in which lysine 3363 in apoB was changed to glutamic acid ($K_{3363} \rightarrow E$), which severely impairs the interaction with artery wall proteoglycans without affecting LDL-receptor-binding activity⁵.

The fifth line expressed 6-GBSM (six glycosaminoglycan binding sites mutated) LDL, created by mutating the six carboxy-terminal of the eight glycosaminoglycan-binding sequences identified in delipidated apoB100⁷⁻⁹. Although residues 3359–3369 are the primary site for the interaction of native LDL with proteoglycans⁵, modification of the LDL might expose other proteoglycan-binding sites. In all constructs, a leucine was substituted for glutamine-2153 to prevent the formation of apoB48 (ref. 10). This was important because proteoglycan binding site(s) other than those on normal LDL are exposed and physiologically important on apoB48 LDL¹¹.

Next, we clarified whether proteoglycan-binding-defective LDL undergo normal non-receptor-mediated transcytosis across the endothelium^{1,12} and compared their subendothelial retention *in vivo* with that of control LDL. $R,K_{3359-3369} \rightarrow S,A$ and recombinant control LDL were labelled with ¹²⁵I and injected into the tail vein of 8-week-old mice. After 20 min or 72 h, the mice were perfusion fixed, and the aortas were dissected and measured for radioactivity. Analysis of aortas 20 min after injection of 5×10^8 c.p.m. of ¹²⁵I-LDL showed no difference in the arterial permeability of the $R,K_{3359-3369} \rightarrow S,A$ LDL and the control LDL (not shown). However, 72 h after injection of 1×10^7 c.p.m. of ¹²⁵I-LDL, more radioactivity was present in the aortas of mice injected with control LDL than of those injected with proteoglycan-binding-defective LDL (263 ± 117 versus 701 ± 324 c.p.m.; mean $\pm$ s.d.; $n = 3$ per group). Thus, proteoglycan-binding-defective LDL are retained less efficiently than recombinant control LDL in the artery wall *in vivo*.

To determine if elevated levels of proteoglycan-binding-defective LDL would be less atherogenic than similar levels of wild-type recombinant LDL, transgenic mice expressing the five types of human recombinant LDL were fed the atherogenic diet for 20 weeks. The lipoprotein profiles (Fig. 1) and the cholesterol levels were similar in all groups (Table 2), except for the mice expressing $R,K_{3359-3369} \rightarrow S,A$ LDL and particularly those expressing

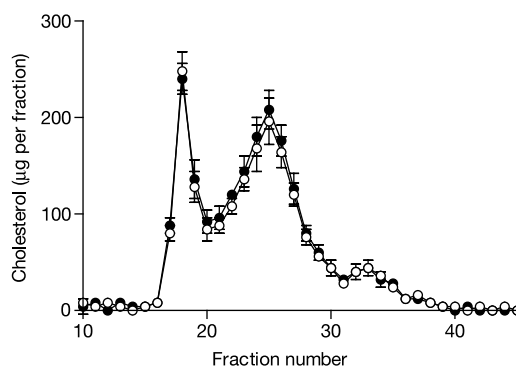


Figure 1 Distribution of cholesterol. Data was assessed by Superose 6 chromatography of pooled plasma from five female human-apoB transgenic mice expressing recombinant control low-density lipoproteins (LDL) (filled circles) or 6-GBSM LDL (open circles) after 20 weeks on an atherogenic diet. Very-low-density lipoproteins correspond to fractions 16–20, LDL to fractions 21–29, and high-density lipoproteins to fractions 30–36.

$K_{3363} \rightarrow E$ LDL, which had a lower plasma LDL-cholesterol level owing to lower transgene copy number.

After 20 weeks, 231 mice were perfusion-fixed. Seven aortas were not analysed for technical reasons. The remaining 224 aortas were analysed with the *en face* procedure¹³. The extent of atherosclerosis correlated with the plasma concentration of human apoB100 in all groups (Fig. 2). However, transgenic mice expressing proteoglycan-binding-defective LDL ($R,K_{3359-3369} \rightarrow S,A$, 6-GBSM, and $K_{3363} \rightarrow E$ LDL) had significantly less atherosclerosis than mice expressing wild-type control or $W_{4369} \rightarrow Y$ LDL (Table 3) ($P < 0.01$). Nontransgenic littermate controls ($n = 5$ per group) had essentially no atherosclerosis (not shown). However, plasma concentrations of human apoB100 were lower in mice expressing $R,K_{3359-3369} \rightarrow S,A$ LDL and $K_{3363} \rightarrow E$ LDL than in those expressing recombinant control LDL, so $W_{4369} \rightarrow Y$ LDL or 6-GBSM LDL (Table 2). We also calculated 95% Boole–Bonferroni simultaneous confidence intervals (99% individual confidence degrees) for each of the five regression slopes, individually. The intervals for the regression slopes of recombinant control LDL and $W_{4369} \rightarrow Y$ LDL did not overlap with those of the three groups of proteoglycan-binding-defective LDL. This finding confirms the pairwise differences in the amount of atherosclerosis between the groups.

These results strongly indicate that proteoglycan-binding-defective LDL have a greatly reduced atherogenic potential and provide direct experimental evidence that direct binding of LDL to artery wall proteoglycans is a key step in atherogenesis.

To verify that the differences in atherogenicity were due solely to different affinities for arterial proteoglycans, we performed several control experiments. First, we analysed the formation of conjugated dienes in $R,K_{3359-3369} \rightarrow S,A$ LDL and recombinant control LDL after copper-stimulated oxidation¹⁴. The lag phase for the formation of conjugated dienes in $R,K_{3359-3369} \rightarrow S,A$ LDL and recombinant control LDL was 79 ± 6 and 74 ± 8 min, respectively, and the

Table 2 Lipid, lipoprotein and apoB measurements in transgenic mice expressing recombinant LDL

	Control LDL	$W_{4369} \rightarrow Y$ LDL	$R,K_{3359-3369} \rightarrow S,A$ LDL	$K_{3363} \rightarrow E$ LDL	6-GBSM LDL
Total cholesterol	8.49 ± 0.97	8.43 ± 1.02	$7.32 \pm 0.92^*$	$5.19 \pm 0.64^*$	9.17 ± 0.90
HDL cholesterol	0.78 ± 0.06	0.76 ± 0.05	0.77 ± 0.04	0.79 ± 0.03	0.75 ± 0.05
Triglyceride	0.18 ± 0.016	0.17 ± 0.014	0.18 ± 0.015	0.19 ± 0.020	0.18 ± 0.013
Human apoB100	$2,334 \pm 59$	$2,326 \pm 67$	$1,935 \pm 55^*$	$1,424 \pm 41^*$	$2,518 \pm 61$

Plasma lipid and lipoprotein (mmol l⁻¹) and apoB levels (μg ml⁻¹) in female mice after 20 weeks on an atherogenic diet. Values are mean $\pm$ s.e.m. ($n = 20$ per group, except for human apoB100, where $n = 40$ per group).

* $P < 0.001$ versus control LDL.

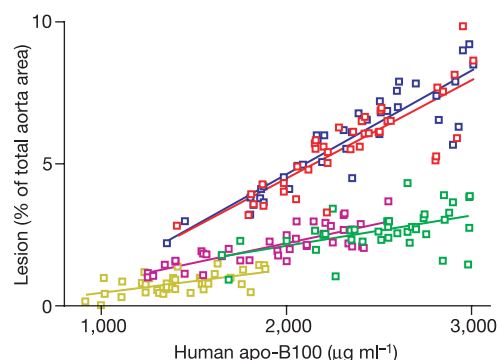


Figure 2 Effect on aorta of an atherogenic diet in transgenic mice. The data shows the correlation between the percentage of total aortic surface area covered by lesions and the plasma concentration of human apoB100 in transgenic mice fed an atherogenic diet for 20 weeks. Recombinant control LDL (red) and $W_{4369} \rightarrow Y$ LDL (blue), both with normal proteoglycan binding. Proteoglycan-binding-defective $R_{K3359-3369} \rightarrow S,A$ LDL (pink), $K_{3363} \rightarrow E$ LDL (light green), and 6-GBSM LDL (dark green). The percentage of total aortic surface area covered by lesions in mice expressing the recombinant LDL were 5.7 ± 1.5 , 5.8 ± 1.7 , 2.1 ± 0.66 , 0.81 ± 0.36 or 2.7 ± 0.72 , respectively (mean $\pm$ s.d.).

maximal rate of conjugated dienes formed was 6.1 ± 0.5 and 5.8 ± 0.4 molecules $\text{min}^{-1} \times \text{LDL}$ particle, respectively (mean $\pm$ s.d.; $n = 3$). Thus, proteoglycan-binding-defective LDL were as susceptible to oxidation as recombinant control LDL.

Next, we incubated minimally oxidized LDL with human and mouse macrophages and analysed the inflammatory response. After incubation with human monocyte-derived macrophages for 24 h, the TNF α concentrations in cell-culture media containing minimally oxidized $R_{K3359-3369} \rightarrow S,A$ LDL or recombinant control LDL were not significantly different (42.2 ± 12.5 and 37.2 ± 10.3 pg per ml medium; mean of duplicate measurements of macrophages from two donors analysed separately). Similarly, after incubation with minimally oxidized $R_{K3359-3369} \rightarrow S,A$ LDL or recombinant control LDL for 24 h, bone-marrow derived macrophages from transgenic mice expressing an NF κ B-luciferase reporter gene showed no differences in luciferase activity (mean $\pm$ s.d., 435 ± 36 and 458 ± 64 relative light units per $10 \mu\text{g}$ protein, respectively; $n = 4$). The concentrations of thiobarbituric-acid-reactive substances¹⁵ (TBARS), lipid hydroperoxide (LPO), and hydrogen peroxide were the same in the minimally oxidized $R_{K3359-3369} \rightarrow S,A$ LDL and the minimally oxidized recombinant control LDL (15 ± 3.5 versus 12.5 ± 4.0 nmol MDA per-mg LDL, 252 ± 45 versus 225 ± 32 nmol per mg LDL, and 2.51 ± 0.85 versus 3.63 ± 1.2 nmol per mg LDL, respectively).

To exclude differences in binding and uptake of proteoglycan-binding-defective LDL and recombinant control LDL in macrophages, LDL were radiolabelled with ¹²⁵I and incubated with human monocyte-derived macrophages for 6 or 24 h. There were no differences in the binding or uptake of recombinant control or proteoglycan-binding-defective LDL (Table 3). We also analysed the

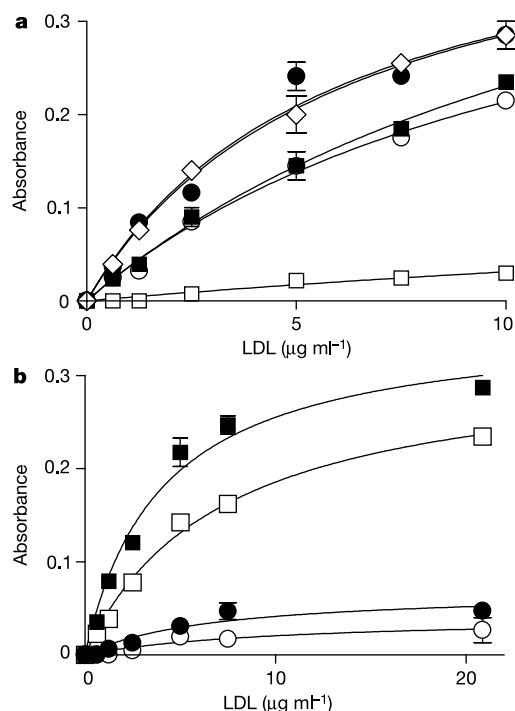


Figure 3 Plate-assay analysis of the ability of recombinant LDL to interact with biglycan. **a**, Recombinant control LDL from $Ldlr^{-/-}$ mice (filled circles) or $Ldlr^{+/+}$ mice (open circles), recombinant $R_{K3359-3369} \rightarrow S,A$ LDL from $Ldlr^{-/-}$ mice (filled squares) or $Ldlr^{+/+}$ mice (open squares), or human plasma LDL (diamonds). **b**, Recombinant $R_{K3359-3369} \rightarrow S,A$ LDL from $Ldlr^{-/-}$ mice (open squares), recombinant $R_{K3359-3369} \rightarrow S,A$ LDL from $Ldlr^{-/-}$ mice with CHD-modified human apoE (open circles), or human plasma LDL (filled squares). CHD-modified human plasma LDL (filled circles) were included as a negative control. The recombinant lipoproteins were isolated from ten mice each, and endogenous apoB was removed by immunoaffinity chromatography.

accumulation of ¹⁴C-cholesterol esters in macrophages incubated with $R_{K3359-3369} \rightarrow S,A$ LDL or recombinant control LDL isolated from mice injected with $[1,2-^{14}\text{C}]$ acetate. The macrophages accumulated similar amounts of ¹⁴C-cholesterol esters after being incubated with radiolabelled $R_{K3359-3369} \rightarrow S,A$ LDL or recombinant control LDL for 24 h ($1,193 \pm 263$ versus $1,274 \pm 382$ c.p.m. per mg cell protein, respectively; mean $\pm$ s.d. of duplicate measurements on macrophages from six donors analysed separately).

Finally, mice expressing $R_{K3359-3369} \rightarrow S,A$ LDL or recombinant control LDL were bred onto the LDL-receptor-null background ($Ldlr^{-/-}$). Because apoE is a ligand for the LDL receptor, the recombinant LDL in $Ldlr^{-/-}$ mice become enriched in apoE. ApoE also facilitates an indirect interaction between LDL and artery wall proteoglycans¹⁶. Therefore, if the decreased atherosclerotic potential of proteoglycan-binding-defective LDL were due solely to its inability to interact with artery wall proteoglycans,

Table 3 Binding and degradation of ¹²⁵I-LDL to human monocyte-derived macrophages

	Membrane-bound LDL		Intracellular LDL		Degraded LDL	
Hours incubated	6	24	6	24	6	24
Human LDL	288 $\pm$ 37	598 $\pm$ 295	709 $\pm$ 59	1,543 $\pm$ 185	1,535 $\pm$ 21	3,857 $\pm$ 1,031
Control LDL	380 $\pm$ 23	787 $\pm$ 549	858 $\pm$ 4	1,822 $\pm$ 344	1,563 $\pm$ 14	5,381 $\pm$ 1,802
$W_{4369} \rightarrow Y$ LDL	282 $\pm$ 63	721 $\pm$ 503	837 $\pm$ 1	1,824 $\pm$ 418	2,033 $\pm$ 28	4,775 $\pm$ 1,156
$R_{K3359-3369} \rightarrow S,A$ LDL	290 $\pm$ 76	845 $\pm$ 720	835 $\pm$ 80	1,762 $\pm$ 761	1,829 $\pm$ 333	4,340 $\pm$ 1,051
$K_{3363} \rightarrow E$ LDL	242 $\pm$ 36	812 $\pm$ 531	781 $\pm$ 75	1,830 $\pm$ 407	1,834 $\pm$ 207	4,882 $\pm$ 1,351
6-GBSM LDL	367 $\pm$ 157	725 $\pm$ 442	798 $\pm$ 122	1,814 $\pm$ 393	1,748 $\pm$ 478	5,068 $\pm$ 1,050

Macrophages were incubated in medium containing ¹²⁵I-labelled LDL for 6 or 24 h. Membrane-bound, cell-associated, and degraded LDL were measured and expressed as ng LDL per mg cell protein. Values are mean $\pm$ s.d. of triplicate measurements on macrophages from five donors (for 24 h) or two donors (for 6 h), analysed separately. No significant differences were found between recombinant proteoglycan-binding LDL and proteoglycan-binding-defective LDL (paired Student's *t*-test).

enrichment in apoE would restore its atherogenicity.

Recombinant wild-type control LDL and R,K₃₃₅₉₋₃₃₆₉ → S,A LDL from *Ldlr*^{+/+} or *Ldlr*^{-/-} mice were subjected to plate-assay analysis with biglycan. R,K₃₃₅₉₋₃₃₆₉ → S,A LDL from *Ldlr*^{-/-} mice (that is, apo-E enriched proteoglycan-binding-defective LDL) interacted normally with biglycan, whereas R,K₃₃₅₉₋₃₃₆₉ → S,A LDL from *Ldlr*^{+/+} mice displayed severely defective binding to biglycan (Fig. 3a). These results were verified in binding experiments with the total proteoglycan fraction from aortic smooth muscle cells (not shown).

To confirm the role of apoE in binding to proteoglycans, we incubated R,K₃₃₅₉₋₃₃₆₉ → S,A LDL from *Ldlr*^{-/-} mice with a 200-fold molar excess of cyclohexanedione (CHD)-modified apoE isolated from human very-low-density lipoprotein¹⁷. In two independent experiments, replacement of the endogenous apoE with CHD-modified human apoE abolished the binding to biglycan (Fig. 3b). We also analysed the binding affinity of R,K₃₃₅₉₋₃₃₆₉ → S,A LDL isolated from *Ldlr*^{-/-} or *ApoE*^{-/-} mice. ApoE-enriched R,K₃₃₅₉₋₃₃₆₉ → S,A LDL from *Ldlr*^{-/-} mice bound biglycan with high affinity (128 ± 21% of control LDL), whereas apoE-depleted R,K₃₃₅₉₋₃₃₆₉ → S,A LDL from *ApoE*^{-/-} mice displayed severely defective binding (6 ± 3% of control LDL).

To determine whether apoE enrichment increases the atherogenicity of proteoglycan-binding-defective R,K₃₃₅₉₋₃₃₆₉ → S,A LDL, we performed atherosclerosis studies in 6-month-old mice. *Ldlr*^{-/-} mice expressing R,K₃₃₅₉₋₃₃₆₉ → S,A LDL or recombinant control LDL developed the same amount of atherosclerosis in both the proximal aortic root (15.8 ± 6.2 and 14.0 ± 4.6% fraction area of lesion, respectively) and distal aorta (4.54 ± 1.31 and 3.94 ± 0.89% of total aorta area, respectively). Thus, the reduced atherogenicity of proteoglycan-binding-defective LDL is dependent on its lower affinity for artery-wall proteoglycans, and the reduced atherogenicity can be restored by apoE enrichment. The lipoprotein profiles (not shown) and the plasma cholesterol levels were similar in transgenic *Ldlr*^{-/-} mice expressing recombinant control LDL or R,K₃₃₅₉₋₃₃₆₉ → S,A LDL (mean ± s.e.m., 15.5 ± 1.3 and 14.7 ± 1.1 mmol l⁻¹, respectively; *n* = 0).

Recombinant LDL from *Ldlr*^{-/-} mice contained approximately 10-fold more apoE than recombinant LDL from *Ldlr*^{+/+} mice. Likewise, the atherogenic diet duplicated the amount of apoE on the recombinant LDL in *Ldlr*^{+/+} mice. The explanation for this is probably a diet-induced down-regulation of LDL receptors. Therefore, proteoglycan-binding-defective LDL from *Ldlr*^{+/+} mice displayed significantly more proteoglycan-binding activity when mice were fed the atherogenic versus the chow (normal food) diet (29 ± 6% and 9 ± 5% of control LDL, respectively). Recombinant control LDL and R,K₃₃₅₉₋₃₃₆₉ → S,A LDL showed identical apoE enrichment after high-fat feeding, and when bred into *Ldlr*^{-/-} mice.

Mouse LDL often contain apoE, but apoB100 is the sole apolipoprotein on human LDL. Thus, bridging molecules are probably less important than a direct interaction between apoB100 and proteoglycans for subendothelial retention of atherogenic lipoproteins in humans. Retained lipoproteins can directly or indirectly provoke all known features of early lesions and, by stimulating local synthesis of proteoglycans, can accelerate further retention and aggregation³. Thus, atherosclerosis is initiated by subendothelial retention of atherogenic lipoproteins. □

Methods

Human apo-B transgenic mice

The human apoB transgenic mice were back-crossed for three generations with C57BL/6 mice or for five generations with *Ldlr*^{-/-} mice on C57BL/6 background (Jackson Laboratory). Mice expressing an NFκB-luciferase reporter gene were a gift from H. Carlsen and R. Blomhoff.

Plasma lipid and lipoprotein measurements

The plasma concentrations of human apoB100 were measured by enzyme-linked

immunosorbent assay (ELISA) using the antihuman apoB antibody 1D1¹⁸ and a horseradish peroxidase (HRP)-conjugated polyclonal antihuman apoB antibody (The Binding Site). The distribution of lipids within the plasma lipoprotein fractions was assessed by fast-performance liquid chromatography (FPLC) gel filtration using a Superose 6 HR 10/30 column (Pharmacia)¹⁹.

Plate-assay analysis with biglycan

Maxisorp immunoplates (NUNC) were coated with biglycan (10 µg ml⁻¹) in HEPES-buffered saline (HBS) (20 mM HEPES, 150 mM NaCl, pH 7.4) overnight at room temperature (RT, 22 °C), and blocked with HBS with 1% bovine serum albumin (BSA) for 1 h at RT. The samples of LDL in HBS buffer with 2 mM CaCl₂ and 2 mM MgCl₂ were added to the wells, and incubated for 1 h at RT. The plates were then incubated with the same buffer supplemented with lipoprotein-deficient serum (diluted 1:50) for 30 min. To each well, 100 µl of a HRP-conjugated polyclonal antibody against human apoB (The Binding Site) (diluted 1:750 in HBS with 0.1% BSA and 0.02% Tween 20) was added and incubated at RT for 1.5 h. Finally, 100 µl of Turbo TMB-ELISA (Pierce) substrate was added and incubated for 5 min.

Analysis of atherosclerotic lesions in the proximal aortic root

The heart and 1 mm of the thoracic aorta were embedded in OCT Tissue-Tec medium (Histolab), frozen in dry ice and isopentane, cut into 10-µm-thick cross-sections, and stained in 0.5% Oil Red O². The fraction area of lesion was calculated by dividing the surface of the lesion by the surface of the vessel. This corrects for errors caused by oblique sections²⁰.

Competition of mouse apoE with chemically modified human apoE

Recombinant R,K₃₃₅₉₋₃₃₆₉ → S,A LDL was isolated⁷ and incubated with a 200-fold molar excess of CHD-modified human very-low-density lipoproteins²¹. The free apoE was removed from the recombinant LDL by ultracentrifugation⁷ (*d* = 1.063 g ml⁻¹). The top 1 ml was recovered and separated on a Superose 6 HR 10/30 column. The CHD-modified samples were incubated with an equal volume of 1 M hydroxylamine, 0.3 M mannitol, pH 7.0, at 37 °C for 16 h to reverse the modification of the arginine residues before gel electrophoresis.

Monitoring LDL oxidation

LPO was quantitated with the Lipohydrox kit (Wak-Chemie Medical, Germany). The H₂O₂ content was measured with the Bioxytech H₂O₂-560 assay (Oxis International, USA).

Macrophage preparation

Human monocyte-derived macrophages, prepared separately from buffy coats (lymphocyte fraction of whole blood) from five donors and obtained from Swedish blood banks, were isolated as described²².

Uptake and degradation of LDL

Human LDL and recombinant mouse LDL were labelled with ¹²⁵I (Amersham International) to a specific activity of 46–54 c.p.m. ng⁻¹ (ref. 23). Macrophages were incubated in medium with ¹²⁵I-labelled LDL (50 µg ml⁻¹) for 6 or 24 h. The membrane-bound, the cell-associated and the degraded LDL were determined²⁴, and expressed as ng LDL per mg cell protein. In another experiment, mice expressing recombinant control LDL or R,K₃₃₅₉₋₃₃₆₉ → S,A LDL were injected intravenously with 80 µCi [1,2-¹⁴C]acetate (Amersham International) 3, 5 and 10 days before the mice were sacrificed and the recombinant LDL isolated⁷. Macrophages were incubated with ¹⁴C-labelled LDL (50 µg ml⁻¹) for 24 h. Lipids were extracted from the cells²⁵, and the radioactivity in ¹⁴C-cholesterol esters was measured after separation by thin-layer chromatography (Silica Gel G) in hexane-diethylether-glacial acetic acid (80:20:1 v/v).

Measurement of TNFα secretion in human macrophages

Human macrophages were cultured for 6 days in RPMI 1640 medium (Bio Whittaker) with 10% human serum and 10% FCS. Macrophages were then incubated for 24 h in serum-free medium containing human or recombinant minimally modified LDL (100 µg ml⁻¹). The medium concentration of TNFα; was measured by high-sensitivity ELISA for TNFα (R&D Systems) and correlated to cell protein content.

Luciferase measurement in bone-marrow-derived mouse macrophages

Bone-marrow-derived mouse macrophages were isolated from the femurs of female NFκB-luciferase reporter mice²⁶ and seeded at 1 × 10⁶ cells ml⁻¹ in RPMI 1640 containing 10% fetal bovine serum (FBS) and 10% L-cell conditioned medium²⁷. Nonadherent cells were removed 24 h later, resuspended at the same density in fresh medium, and cultured for 6 days until confluence was reached. The mouse macrophages were then incubated for 24 h in serum-free medium containing minimally modified human or recombinant LDL (100 µg ml⁻¹). Luciferase activity (Luciferase Assay System, Promega) was determined and normalized for protein content.

Retention of LDL *in vivo*

Recombinant control LDL and R,K₃₃₅₉₋₃₃₆₉ → S,A LDL were isolated⁷ and radiolabelled with ¹²⁵I using Iodogen (Pierce)²⁸. The specific activity of the recombinant control ¹²⁵I-LDL and the R,K₃₃₅₉₋₃₃₆₉ → S,A ¹²⁵I-LDL was 300 c.p.m. ng⁻¹.

Statistical analysis

Differences in atherosclerosis and lipid levels were assessed by analysis of variance (ANOVA) on ranks and multiple comparison analysis.

Received 6 December 2001; accepted 20 February 2002; doi:10.1038/nature00804.

- Ross, R. Cell biology of atherosclerosis. *Annu. Rev. Physiol.* **57**, 791–804 (1995).
- Glass, C. K. & Witztum, J. L. Atherosclerosis. The Road Ahead. *Cell* **104**, 503–516 (2001).
- Williams, K. J. & Tabas, I. The response-to-retention hypothesis of early atherogenesis. *Arterioscler. Thromb. Vasc. Biol.* **15**, 551–561 (1995).
- Srinivasan, S. R. *et al.* Low density lipoprotein retention by aortic tissue. Contribution of extracellular matrix. *Atherosclerosis* **62**, 201–208 (1986).
- Boren, J. *et al.* Identification of the principal proteoglycan-binding site in LDL. A single-point mutation in apo-B100 severely affects proteoglycan interaction without affecting LDL receptor binding. *J. Clin. Invest.* **101**, 2658–2664 (1998).
- Boren, J. *et al.* Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100. *J. Clin. Invest.* **101**, 1084–1093 (1998).
- Weisgraber, K. H. & Rall, S. C. Jr Human apolipoprotein B-100 heparin-binding sites. *J. Biol. Chem.* **262**, 11097–11103 (1987).
- Hirose, N., Blankenship, D. T., Krivanek, M. A., Jackson, R. L. & Cardin, A. D. Isolation and characterization of four heparin-binding cyanogen bromide peptides of human plasma apolipoprotein B. *Biochemistry* **26**, 5505–5512 (1987).
- Camejo, G., Olofsson, S. O., Lopez, F., Carlsson, P. & Bondjers, G. Identification of Apo B-100 segments mediating the interaction of low density lipoproteins with arterial proteoglycans. *Arteriosclerosis* **8**, 368–377 (1988).
- Yao, Z. *et al.* Elimination of apolipoprotein B48 formation in rat hepatoma cell lines transfected with mutant human apolipoprotein B cDNA constructs. *J. Biol. Chem.* **267**, 1175–1182 (1992).
- Goldberg, I. J. *et al.* The NH2-terminal region of apolipoprotein B is sufficient for lipoprotein association with glycosaminoglycans. *J. Biol. Chem.* **273**, 35355–35361 (1998).
- Simionescu, M. & Simionescu, N. Endothelial transport of macromolecules: transcytosis and endocytosis. A look from cell biology. *Cell Biol. Rev.* **25**, 5–78 (1991).
- Tangirala, R. K., Rubin, E. M. & Palinski, W. Quantitation of atherosclerosis in murine models: Correlation between lesions in the aortic origin and in the entire aorta, and differences in the extent of lesions between sexes in LDL receptor-deficient and apolipoprotein E-deficient mice. *J. Lipid Res.* **36**, 2320–2328 (1995).
- Puhl, H., Waeg, G. & Esterbauer, H. Methods to determine oxidation of low-density lipoproteins. *Methods Enzymol.* **233**, 425–441 (1994).
- Yagi, K. A simple fluorometric assay for lipoperoxide in blood plasma. *Biochem. Med.* **15**, 212–216 (1976).
- Ji, Z. S., Pitas, R. E. & Mahley, R. W. Differential cellular accumulation/retention of apolipoprotein E mediated by cell surface heparan sulfate proteoglycans. Apolipoproteins E3 and E2 greater than E4. *J. Biol. Chem.* **273**, 13452–13460 (1998).
- Brisette, L., Roach, P. D. & Noel, S. P. The effects of liposome-reconstituted apolipoproteins on the binding of rat intermediate density lipoproteins to rat liver membranes. *J. Biol. Chem.* **261**, 11631–11638 (1986).
- Milne, R. W., Theolis, R. Jr, Verdery, R. B. & Marcel, Y. L. Characterization of monoclonal antibodies against human low density lipoprotein. *Arteriosclerosis* **3**, 23–30 (1983).
- Purcell-Huynh, D. A. *et al.* Transgenic mice expressing high levels of human apolipoprotein B develop severe atherosclerotic lesions in response to a high-fat diet. *J. Clin. Invest.* **95**, 2246–2257 (1995).
- Nicoletti, A., Kaveri, S., Caligiuri, G., Bariety, J. & Hansson, G. K. Immunoglobulin treatment reduces atherosclerosis in apo E knockout mice. *J. Clin. Invest.* **102**, 910–918 (1998).
- Mahley, R. W. *et al.* Inhibition of lipoprotein binding to cell surface receptors of fibroblasts following selective modification of arginyl residues in arginine-rich and B apoproteins. *J. Biol. Chem.* **252**, 7279–7287 (1977).
- Ohlsson, B. G. *et al.* Oxidized low density lipoprotein inhibits lipopolysaccharide-induced binding of nuclear factor- κ B to DNA and the subsequent expression of tumour necrosis factor- α and interleukin-1 β in macrophages. *J. Clin. Invest.* **98**, 78–89 (1996).
- McFarlane, A. S. Efficient trace-labelling of proteins with iodine. *Nature* **182**, 53 (1958).
- Hurt-Camejo, E. *et al.* Effect of arterial proteoglycans and glycosaminoglycans on low density lipoprotein oxidation and its uptake by human macrophages and arterial smooth muscle cells. *Arterioscler. Thromb.* **12**, 569–583 (1992).
- Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
- Randle, D. H., Zindy, F., Sherr, C. J. & Roussel, M. F. Differential effects of p19^{Arf} and p16^{Ink4a} loss on senescence of murine bone marrow-derived preB cells and macrophages. *Proc. Natl Acad. Sci. USA* **98**, 9654–9659 (2001).
- Stanley, E. R. The macrophage colony-stimulating factor, CSF-1. *Methods Enzymol.* **116**, 564–587 (1985).
- Schwenke, D. C. Gender differences in intima-media permeability to low-density lipoprotein at atherosclerosis-prone aortic sites in rabbits. Lack of effect of 17 β -estradiol. *Arterioscler. Thromb. Vasc. Biol.* **17**, 2150–2157 (1997).

Acknowledgements

We thank L. Lindgren, C. Ullström and A. Lidell for technical assistance, O. Nerman and K. Wiklander for statistical analysis, D. Schwenke for advice with retention studies. K. Weisgraber for comments on the manuscript, and S. Ordway and G. Howard for editorial assistance. This work was supported by the Swedish Medical Research Council, The Swedish Foundation for Strategic Research, The Swedish Heart–Lung Foundation, and in part by a National Institutes of Health grant.

Competing interests statement

The authors declare that they have no competing financial interests

Correspondence and requests for materials should be addressed to J.B. (e-mail: jan.boren@wlab.gu.se).

T-box gene *tbx5* is essential for formation of the pectoral limb bud

Dae-gwon Ahn^{*†}, Matthew J. Kourakis^{*†}, Laurel A. Rohde^{*†}, Lee M. Silver[†] & Robert K. Ho^{*†}

^{*} Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA.

[†] Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

The T-box genes *Tbx4* and *Tbx5* have been shown to have key functions in the specification of the identity of the vertebrate forelimb (*Tbx5*) and hindlimb (*Tbx4*)^{1,2}. Here we show that in zebrafish, *Tbx5* has an additional early function that precedes the formation of the limb bud itself. Functional knockdown of zebrafish *tbx5* through the use of an antisense oligonucleotide resulted in a failure to initiate fin bud formation, leading to the complete loss of pectoral fins. The function of the *tbx5* gene in the development of zebrafish forelimbs seems to involve the directed migration of individual lateral-plate mesodermal cells into the future limb-bud-producing region. The primary defect seen in the *tbx5*-knockdown phenotype is similar to the primary defects described in known T-box-gene mutants such as the *spadetail* mutant of zebrafish^{3,4} and the *Brachyury* mutant of the mouse⁵, which both similarly exhibit an altered migration of mesodermal cells. A common function for many of the T-box genes might therefore be in mediating the proper migration and/or changes in adhesive properties of early embryonic cells.

The formation of vertebrate limbs involves a complex series of morphogenetic events, including the specification of limb fields within the lateral-plate mesoderm, induction of the limb buds at appropriate axial levels, and the initiation and patterning of distal limb outgrowth^{6,7}. During early stages of vertebrate limb morphogenesis, *Tbx5* is strongly expressed within the forelimb buds of a variety of vertebrate species^{8–13}. Recent misexpression studies have shown that this gene and a closely related gene, *Tbx4*, which is expressed within the hindlimb bud, are crucial in the determination of limb identity and the regulation of limb outgrowth^{1,2}. However, the initiation of *Tbx5* expression within the anterior lateral-plate mesoderm, which supplies the forelimb progenitor cells, precedes the emergence of visible forelimb buds^{8–11,13}, indicating that this gene might have additional early functions in forelimb development. To investigate the possibility of an earlier function for the *Tbx5* gene during vertebrate forelimb development, we generated a knockdown phenotype in zebrafish, using antisense oligonucleotides containing morpholino moieties in their backbones. These ‘morpholino’ oligonucleotides are thought to exert their inhibitory effects through physical blocking of the translational initiation of target messenger RNAs, and also have been shown to exhibit a low toxicity and high specificity in a variety of *in vivo* systems¹⁴, including the fertilized eggs of zebrafish¹⁵.

We designed two different morpholino oligonucleotides for the *tbx5* gene of zebrafish, one recognizing the first 25 bases of the coding sequence, the other targeting a sequence of a similar size but located within the 5′ untranslated region (UTR), ten nucleotides upstream of the initiation codon. The effectiveness of these oligonucleotides in inhibiting the translation of target mRNAs was first examined by an assay *in vivo* using chimaeric mRNAs in which the coding sequence of the gene encoding enhanced green fluorescent protein replaced the coding sequence of *tbx5* in frame after the first 27 bases. Both oligonucleotides were able to block the translation of the green fluorescent protein when injected together with the test mRNA into the early-stage embryo, whereas the control oligonucleotide designed for the 5′ UTR sequence of the zebrafish *tbx4* gene

was not, indicating that both oligonucleotides are capable of recognizing their targets in a sequence-specific manner under *in vivo* conditions (see Supplementary Information). In addition, when injected into early-stage embryos, both morpholino oligonucleotides generated indistinguishable phenotypic outcomes for the pectoral fins that were not seen in embryos injected with the control oligonucleotide (Fig. 1), again confirming the specific action of these oligonucleotides.

Introduction of 5 ng or more of either morpholino oligonucleotide by microinjection into the embryo at the one-cell or two-cell stage resulted in a complete lack of pectoral fins in all of the injected embryos at 3 days of development ($n = 500$; Fig. 1a, b). Fish examined at 4 days of development were found to lack all of the skeletal elements of the pectoral fins lying distal to the cleithrum, including the scapulocoracoid and the endoskeletal disc (Fig. 1c–f),

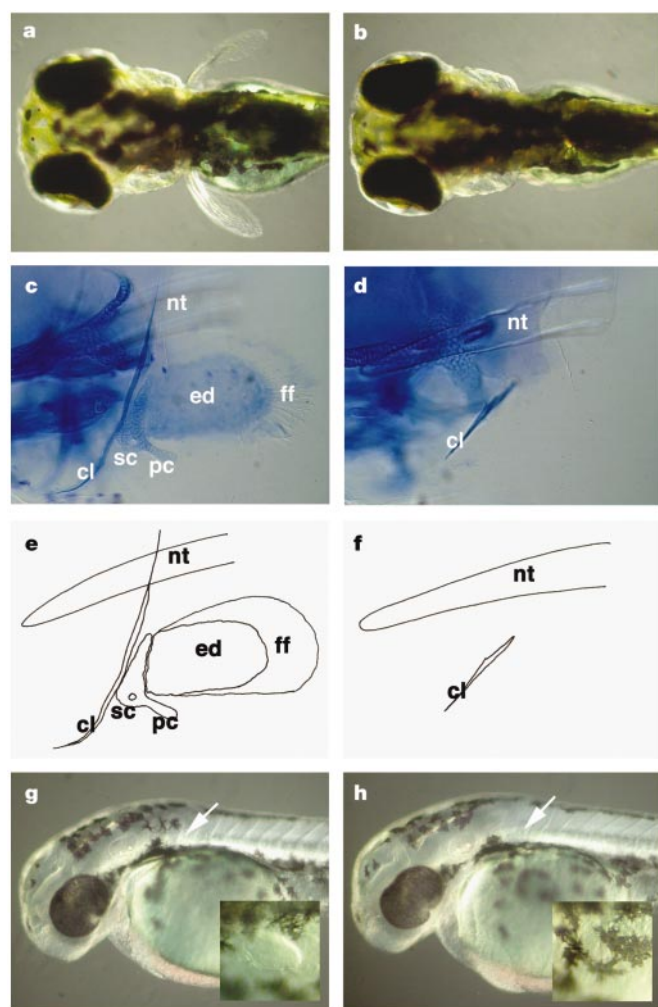


Figure 1 Defects in pectoral fin formation in zebrafish embryos injected with anti-*tbx5* morpholino oligonucleotides. **a, c, e, g**, Embryos injected with the control oligonucleotide (5 ng per embryo). **b, d, f, h**, Embryos injected with an anti-*tbx5* oligonucleotide (5 ng per embryo). **a, b**, Dorsal view of 3-day-old embryos; **c, d**, lateral views of pectoral fin skeletons of 4-day-old larval fish stained with Alcian blue¹⁶; **e, f**, line drawings of the notochord and the major skeletal elements of the pectoral fins shown in **c** and **d**; **g, h**, lateral views of 36-hour-old embryos. Insets show the dorsal views of the prospective pectoral fin regions on the left side (marked by arrows) at a high magnification. Note that at this concentration no gross abnormality was noted in either group of embryos except for the complete lack of pectoral fins in embryos injected with an anti-*tbx5* morpholino oligonucleotide. nt, notochord; cl, cleithrum; sc, scapulocoracoid; pc, postcoracoid process; ed, endoskeletal disc; ff, fin fold.

which give rise to the girdle and the endoskeletal supports of the pectoral fins of the juvenile and adult fish¹⁶. These results indicate that the inhibition of translation of *tbx5* mRNAs by morpholino oligonucleotides has led to the loss of all pectoral limb-bud-derived structures. Our observations also indicate that the knockdown of *tbx5* function might directly affect limb-bud formation at a very early stage, as a pectoral fin bud was never visibly apparent at any stage of development in *tbx5* morpholino-injected embryos (Fig. 1g, h).

In both tetrapod and fish embryos, limb development begins with the formation of a local condensation of mesenchyme cells within the prospective limb field. Signals from the mesenchyme then instruct the overlying ectoderm to form the apical ectodermal ridge (AER), which in turn has a function in initiating and maintaining a cascade of events, including the specification of the zone of polarizing activity (ZPA) in the posterior part of the limb-bud mesenchyme^{6,7}. Members of the *Hox* cluster genes are also expressed very early in the limb bud, some of which appear prior to and independently of the establishment of the ZPA and/or AER^{17–20}. In *tbx5* morpholino-treated embryos, expression of known early markers of fin/limb-bud formation, such as the AER-specific marker *dlx2*, the ZPA-specific marker *shh* and the fin/limb-specific *hoxa* and *hoxd* cluster genes, is never observed (Fig. 2), indicating that, in the absence of *tbx5* function, several of the events characterizing early periods of pectoral fin development were not initiated. These results are consistent with the normal onset of expression of *tbx5* within the anterior lateral-plate mesoderm before the initiation of pectoral fin-bud formation in zebrafish^{13,21}, indicating that *tbx5* might begin to function very early during pectoral fin development, possibly even before the initial formation of the mesenchymal core of the fin bud.

The normal morphogenesis of the *tbx5*-positive cells in the lateral-plate mesoderm of a control embryo is shown in Fig. 3a, c, e, g. In the 10-somite-stage embryo (14 h post fertilization (hpf); Fig. 3a), the *tbx5*-positive lateral-plate cells are arranged in two bilateral stripes extending from the level of the posterior midbrain to the second somite in the trunk. By the 20-somite stage (19 hpf; Fig. 3c), the *tbx5*-expressing cells are no longer arranged in a

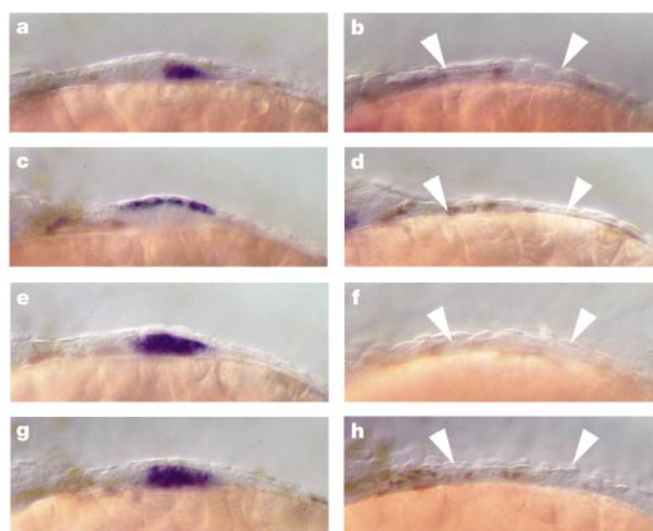


Figure 2 Expression of marker genes in embryos injected with the control oligonucleotide (**a, c, e, g**) (5 ng per embryo) and an anti-*tbx5* morpholino oligonucleotide (**b, d, f, h**) (5 ng per embryo). **a, b**, *shh*; **c, d**, *dlx2*; **e, f**, *hoxa9b*; **g, h**, *hoxd9a*. Embryos at 30 hpf. Arrowheads in **b, d, f** and **h** mark the prospective pectoral fin bud region in *tbx5* morpholino-injected embryos. Note the flat appearance of the pectoral fin region in embryos injected with the anti-*tbx5* morpholino oligonucleotide.

contiguous stripe but are broken up into an anterior group of cells (asterisk), which seems later to contribute to the heart primordia, and a more posterior group of cells (arrow) at the position of the future pectoral fin bud. The pattern of *tbx5* expression within this posterior group of cells further condenses by 24–30 hpf such that most of the *tbx5*-expressing cells become located within the mesenchymal region of the forming pectoral fin buds (Fig. 3e, g).

In embryos injected with the *tbx5* morpholino antisense oligonucleotide, the initiation of *tbx5* expression within the anterior lateral-plate mesoderm (Fig. 3b) as well as the subsequent elaboration of the *tbx5* expression domains into the anterior heart and posterior pectoral fin primordia (Fig. 3d) seem to occur normally, although the overall arrangement of the expressing cells seems somewhat more loosely organized in these embryos than in control embryos. However, by 24 hpf, when the control embryos begin to initiate the formation of pectoral fins as evidenced by the presence of a dense local population of *tbx5*-expressing cells (Fig. 3e, arrow), in the morpholino-injected embryos, the *tbx5*-expressing mesenchyme cells in the prospective fin field remain dispersed (Fig. 3f), and

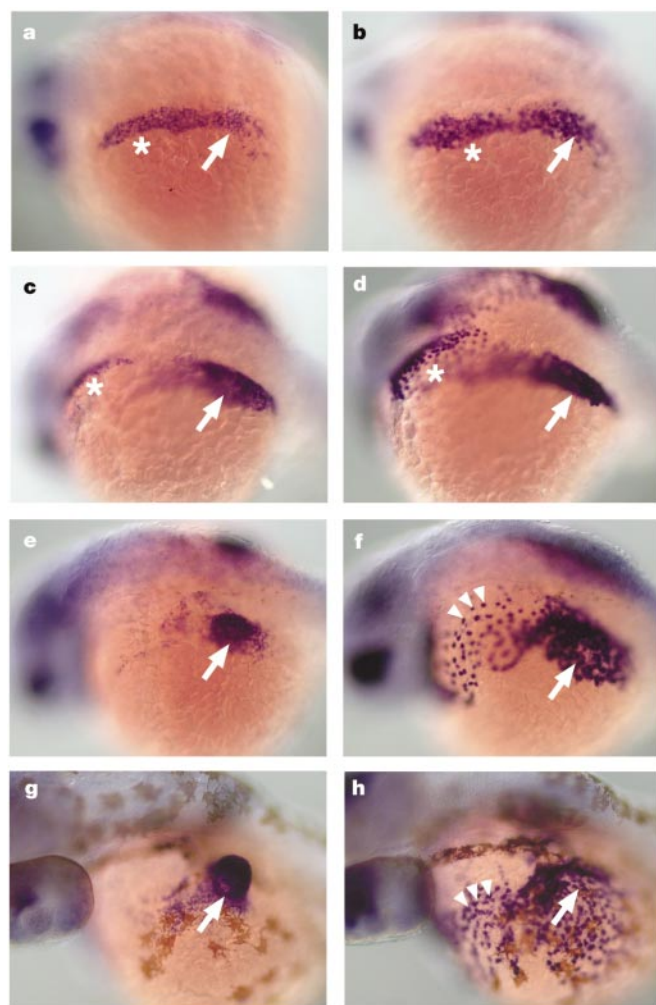


Figure 3 Expression of *tbx5* in embryos injected with the control oligonucleotide (**a, c, e, g**) (5 ng per embryo) and an anti-*tbx5* morpholino oligonucleotide (**b, d, f, h**) (5 ng per embryo). **a, b**, Ten-somite stage; **c, d**, 20-somite stage; **e, f**, 24 hpf; **g, h**, 36 hpf. Oblique dorsal view of the left side of the embryo, with the anterior to the left and dorsal to the top in all panels. Arrows and asterisks mark the pectoral fin and heart primordia, respectively. Small arrowheads in **f** and **h** mark the cells of the pericardium in *tbx5* morpholino-injected embryos showing high levels of *tbx5* expression, which normally express *tbx5* at this stage and earlier, but only weakly (D.A., unpublished observation).

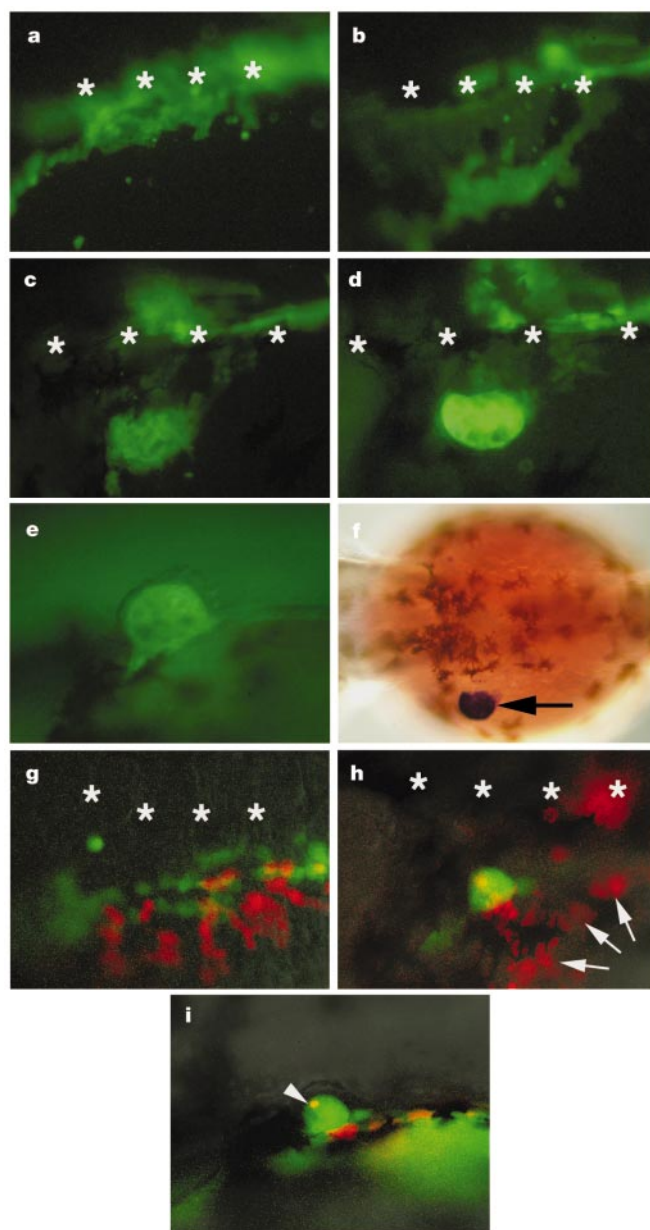


Figure 4 Rescued formation of a pectoral fin bud by transplanted wild-type cells in a *tbx5* morpholino-injected embryo. **a–e**, Images of the prospective pectoral fin fields taken at different time points during the development of a *tbx5* morpholino-injected host carrying transplanted wild-type cells (green). **f**, Dorsal view of the same specimen after whole-mount *in situ* hybridization with the *hoxa9b* probe, a marker for the pectoral fin mesenchyme cells^{17,18}. Note the strong expression of *hoxa9b* within the rescued fin bud (arrow), which is not seen in the contralateral side that did not receive wild-type cells. **a**, Ten-somite stage; **b**, 26-somite stage (22 hpf); **c**, 30 hpf; **d–f**, 36 hpf. Fluorescent (**a–d**), composite (**e**) or white-light (**f**) images. In all rescues, the congregation of the wild-type cells always occurred at the correct axial position for normal pectoral fin outgrowth (in the area adjacent to the second somite). No rescues were seen in host embryos in which wild-type cells were located outside the lateral-plate mesoderm at the level of the first three somites ($n = 63$). **g–i**, Differential contribution of wild-type (green) and *tbx5* morpholino-treated (red) cells to the formation of a pectoral fin bud in a *tbx5* morpholino-injected host. **g**, Ten-somite stage; **h, i**, 36 hpf. Note the sharp boundary separating the wild-type cells (green) from the underlying layer of morpholino-treated cells (red). The single red spot in **i** (arrowhead) is a dead cell fragment, presumably swept into the fin bud by the migration of wild-type cells. In the remaining cases of rescue ($n = 4$), morpholino-injected cells were excluded from the mesenchymal layers of the fin bud. Oblique dorsal (**a–d, g, h**), lateral (**e, i**) and dorsal (**f**) views, with the anterior to the left and dorsal to the top in all panels. Asterisks in **a–d, g, h** mark the anterior and posterior ends of the ventral borders of the first three somites, with the distance between the marks covering the width of each somite ($\sim 50 \mu\text{m}$).

eventually are scattered over the yolk away from the normal position of the pectoral fin bud (Fig. 3h).

The patterns of *tbx5* expression in both control and experimental embryos indicate that one of the functions of the *Tbx5* gene is to mediate the correct movement of lateral-plate mesoderm cells into the future pectoral limb-bud-producing region. In support of this possibility, we found that wild-type cells, if transplanted into the anterior lateral-plate region of *tbx5* morpholino-treated embryos, can rescue pectoral fin buds through a directed migration. As illustrated in Fig. 4a–e, which is a series of photographs taken at various time points from one of the rescued embryos, we observed that fluorescently labelled wild-type cells within a region adjacent to the first three somites (the area that normally contributes cells to the pectoral fin buds (D.A., unpublished observation)) were reproducibly able to undergo cell movements (Fig. 4a, b) and condense

(Fig. 4c, d) to form a pectoral fin bud (Fig. 4e, f) within morpholino-treated host embryos ($n = 7$). Furthermore, when we simultaneously transplanted cells from a wild-type donor embryo and from a morpholino-treated donor embryo into a morpholino-treated host embryo, we found that only the wild-type cells (labelled green in Fig. 4g–i) were capable of forming and populating the rescued fin bud structure, whereas morpholino-treated cells (labelled red in Fig. 4g–i) remained outside the fin bud in a single layer of variably scattered cells around the limb-producing region (Fig. 4h, arrows; $n = 5$). These results indicate that *tbx5* function is required autonomously to allow cells to populate and form the mesenchymal core of the developing pectoral fin bud.

It has been suggested from studies of known genetic mutants in T-box genes, such as the *Brachyury* or *T* mutant in mouse⁵ and the *spadetail* mutant in zebrafish, which carries a defect in the *tbx16* gene⁴, that a general function of many of the T-box family genes might be in mediating the correct migration of cells, perhaps through changes in adhesive properties, to their proper positions in the developing embryo^{3–5,22}. For example, the primary defect described in the *spadetail* mutant is a cell-autonomous failure of marginal mesodermal cells to converge correctly to the dorsal midline during gastrulation³, resulting in an embryo lacking somitic trunk mesoderm. Likewise, deficiencies and abnormal morphogenesis of mesodermal derivatives in *Brachyury*-homozygous mutants have been shown to be due to the inability of mesodermal cells to move correctly through the primitive streak during gastrulation in mouse embryos⁵. The failure of *tbx5* morpholino-treated lateral-plate cells to form mesenchymal aggregates, coupled with the normal behaviours of wild-type cells observed within morpholino-treated host embryos, indicates that the *tbx5* gene in zebrafish embryos might similarly be acting through the regulation of migration and/or adhesive properties of limb precursor cells, either before or during the formation of the pectoral fin bud.

In contrast to the drastic effects of *tbx5* knockdowns on the development of pectoral fins, we saw only subtle and late defects in patterning of the hearts in morpholino-injected embryos, despite the fact that *tbx5* is also prominently expressed in the heart primordia during zebrafish development^{13,21}. In 5-day-old zebrafish, the completion of heart ‘looping’ normally positions the atrium dorsally to the ventricle. However, in *tbx5* morpholino-injected embryos, the heart fails to undergo this late phase of heart looping, leading to the placement of both heart chambers at the same dorsoventral level with the atrium positioned posteriorly to the ventricle (see Supplementary Information). In spite of the failure to complete looping, the beating and unidirectional flow of blood are not significantly impaired in the hearts of a *tbx5*-knockdown fish.

The heart phenotype observed in the *tbx5*-knockdown zebrafish is mild compared with the phenotype reported in mammals, in which the loss or reduction of *Tbx5* function has been shown to lead to much more severe defects, such as hypoplasia of the posterior segment of the heart tissue and abnormalities in cardiac septation²³. The differences in heart phenotypes between *tbx5*-knockdown fish and *Tbx5*-knockout mice could be due to limitations in the morpholino antisense technique that we have used in this study. Alternatively, it is possible that the differences in phenotype might reflect differential requirements for *Tbx5* function in heart development between higher and lower vertebrates.

However, we have found that by using a smaller amount of morpholino oligonucleotides (less than 3 ng per embryo) a series of intermediate phenotypes could be generated for the pectoral fins (Fig. 5a–f), which indicates a possible conservation of *Tbx5* dose dependence in pectoral limb development between fish and tetrapods. We also found that in these instances the reduction in fin size was accompanied by structural defects, such as large tissue gaps, only on the dorsal side of the endoskeletal discs (Fig. 5a, c, e). Similar asymmetries in limb defects have previously been docu-

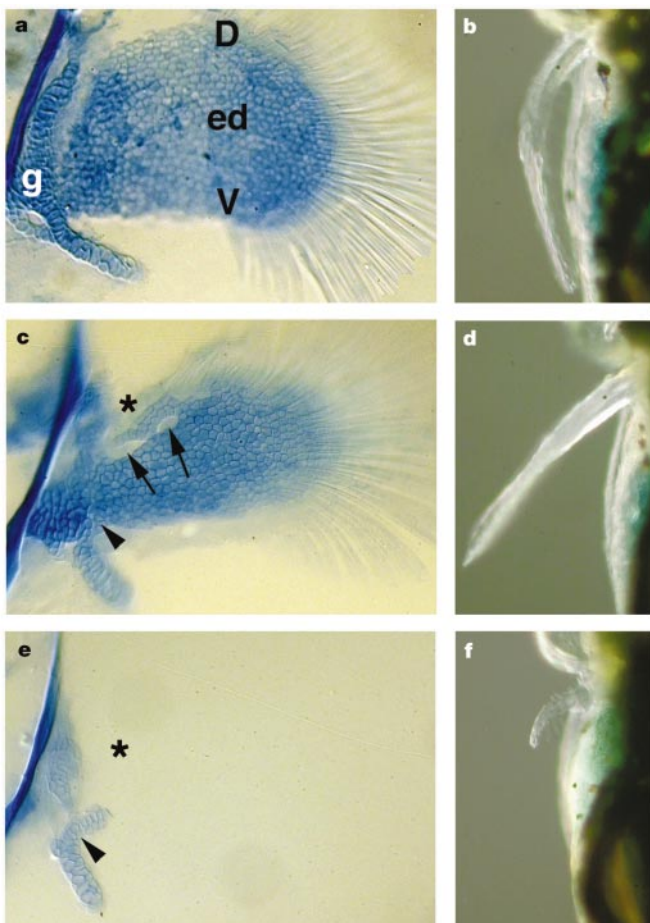


Figure 5 Reduction in pectoral fin size and defects in cartilage formation in 5-day-old larval zebrafish injected with low concentrations of anti-*tbx5* morpholino oligonucleotide (1–3 ng per embryo). **a, b**, A normal pectoral fin from embryos injected with the lowest dose of morpholino (1 ng per embryo). **c–f**, Smaller pectoral fins from embryos injected with a medium dose of *tbx5* morpholino oligonucleotides (2–3 ng per embryo) showing an abnormal development of cartilage structures. These include a partial fusion between the girdle (g) and the endoskeletal disc (ed) (arrowheads in **c** and **e**), which contributes to the abnormal angle of articulation between the pectoral fin and the girdle (as seen in **d**), an incomplete formation (arrows in **c**) or severe reduction of the dorsal portion of the endoskeletal disc (marked by asterisks in **c** and **e**) that accompanies the reduction in disc size, and a malformation (**c**) or splitting (**e**) of the pectoral girdle. Note that, because of the rotation of fin buds during the morphogenesis of pectoral fins in zebrafish, the dorsal portion of larval pectoral fins actually derives from the anterior half of the fin bud¹⁶. D, dorsal; V, ventral.

mented to occur in Holt–Oram syndrome patients, who possess only one good copy of the *Tbx5* gene^{24,25}. At present, however, it is not clear whether such similarities in phenotypes are due to similarities in underlying mechanisms. For instance, here we have reported our findings showing that lateral-plate mesoderm cells in zebrafish seem to migrate extensively during formation of the limb bud. However, in the mouse and chick, *Tbx5*-expressing cells do not seem to undergo similar early migration movements^{8–11}. Future studies on the precise functions of *Tbx5* genes in both higher and lower vertebrates, including fate-mapping work and the isolation of downstream target genes, will be instructive in determining the degree of conservation and divergence of *Tbx5* functions in vertebrate limb development and evolution. □

Methods

Morpholino oligonucleotides were purchased from Gene Tools LLC with the following sequences: control oligonucleotide (anti-*tbx4* 5' UTR), 5'-CCAGAACGCGAGTAATTGTCCACTT-3'; anti-*tbx5* oligonucleotide for the coding sequence, 5'-GAAAGGTGTCTTCACTGTCCGCAT-3'; anti-*tbx5* oligonucleotide for the 5' UTR sequence, 5'-CCTGTACGATGTCTACCGTGAGGC-3'. Solubilization and injection of oligonucleotides were performed as described¹⁵. Unless indicated otherwise, the results presented here are based on the anti-*tbx5* oligonucleotide for the 5' UTR sequence. Alcian blue staining of the cartilages in larval fish and whole-mount *in situ* hybridization on fixed embryos were performed with the procedures described in refs 16 and 26, respectively. Transplantation experiments were performed as described³. In brief, for single transplantation experiments, wild-type donor embryos were first labelled with fluorescein–dextran (relative molecular mass 40,000 (*M_r* 40K)) at the one-cell or two-cell stage. When the donor embryos reached late blastula stage, about 10–20 cells were taken out of each donor embryo and were subsequently introduced into the marginal zones of early gastrula-stage host embryos that had been preinjected with 5 ng anti-*tbx5* oligonucleotide at the one-cell to four-cell stage. Double-transplantation experiments were performed with the same procedure, except for the simultaneous use of both wild-type donor embryos (labelled with *M_r* 40K fluorescein–dextran, green) and *tbx5* morpholino-injected donor embryos (labelled with *M_r* 10K rhodamine–dextran, red).

Received 4 February; accepted 26 March 2002; doi:10.1038/nature00814.

- Takeuchi, J. K. *et al.* *Tbx5* and *Tbx4* genes determine the wing/leg identity of limb buds. *Nature* **398**, 810–814 (1999).
- Rodriguez-Esteban, C. *et al.* The *Tbx* genes *Tbx4* and *Tbx5* regulate limb outgrowth and identity. *Nature* **398**, 814–818 (1999).
- Ho, R. K. & Kane, D. A. Cell-autonomous action of zebrafish *spt-1* mutation in specific mesodermal precursors. *Nature* **348**, 728–730 (1990).
- Griffin, K. J. P., Amacher, S. L., Kimmel, C. B. & Kimmel, D. Molecular identification of *spadetail*: regulation of zebrafish trunk and tail mesoderm formation by *Tbx* genes. *Development* **125**, 3379–3388 (1998).
- Wilson, V., Manson, L., Skarnes, W. C. & Beddington, R. S. The *T* gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development* **121**, 877–886 (1995).
- Tickle, C. & Eichele, G. Vertebrate limb development. *Annu. Rev. Cell Biol.* **10**, 121–152 (1994).
- Cohn, M. J. & Bright, P. E. Molecular control of vertebrate limb development, evolution and congenital malformations. *Cell Tiss. Res.* **296**, 3–17 (1999).
- Gibson-Brown, J. J. *et al.* Evidence of a role for *Tbx* genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. *Mech. Dev.* **56**, 93–101 (1996).
- Ohuchi, H. *et al.* Correlation of wing-leg identity in ectopic FGF-induced chimeric limbs with the differential expression of chick *Tbx5* and *Tbx4*. *Development* **125**, 51–60 (1998).
- Isaac, A. *et al.* *Tbx* genes and limb identity in chick embryo development. *Development* **125**, 1867–1875 (1998).
- Gibson-Brown, J. J., Agulnik, S. I., Silver, L. M., Niswander, L. & Papaioannou, V. E. Involvement of *Tbx* genes *Tbx2–Tbx5* in vertebrate limb specification and development. *Development* **125**, 2499–2509 (1998).
- Takabatake, Y., Takabatake, T. & Takeshima, K. Conserved and divergent expression of *Tbx* genes *Tbx2–Tbx5* in *Xenopus*. *Mech. Dev.* **91**, 433–437 (2000).
- Ruvinsky, I., Oates, A. C., Silver, L. M. & Ho, R. K. The evolution of paired appendages in vertebrates: *Tbx* genes in zebrafish. *Dev. Genes Evol.* **210**, 82–91 (2000).
- Summerton, J. Morpholino anti-sense oligomers: the case for an RNaseH-independent structural type. *Biochim. Biophys. Acta* **1489**, 141–158 (1999).
- Nasevicius, A. & Ekker, S. C. Effective targeted gene 'knockdown' in zebrafish. *Nature Genet.* **26**, 215–220 (2000).
- Grandel, H. & Schulte-Merker, S. The development of the paired fins in the zebrafish (*Danio rerio*). *Mech. Dev.* **79**, 99–120 (1998).
- Neumann, C. J., Grandel, H., Gaffield, W., Schulte-Merker, S. & Nüsslein-Volhard, C. Transient establishment of anteroposterior polarity in the zebrafish pectoral fin bud in the absence of *sonic hedgehog* activity. *Development* **126**, 4817–4826 (1999).
- Grandel, H., Draper, B. W. & Schulte-Merker, S. *dackel* acts in the ectoderm of the zebrafish pectoral fin bud to maintain AER signalling. *Development* **127**, 4169–4178 (2000).
- Ros, M. A. *et al.* The limb field mesoderm determines initial limb bud anteroposterior asymmetry and budding independent of *sonic hedgehog* or apical ectodermal gene expressions. *Development* **122**, 2319–2330 (1996).
- Nelson, C. E. *et al.* Analysis of *Hox* gene expression in the chick limb bud. *Development* **122**, 1449–1466 (1996).
- Begemann, G. & Ingham, P. W. Developmental regulation of *Tbx5* in zebrafish embryogenesis. *Mech. Dev.* **90**, 299–304 (2000).

- Russ, A. P. *et al.* *Eomesodermin* is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95–99 (2000).
- Bruneau, B. G. *et al.* A murine model of Holt–Oram syndrome defines roles of the *Tbx* transcription factor *Tbx5* in cardiogenesis and disease. *Cell* **106**, 709–721 (2001).
- Li, Q. Y. *et al.* Holt–Oram syndrome is caused by mutation in *TBX5*, a member of the Brachyury (*T*) gene family. *Nature Genet.* **15**, 21–29 (1997).
- Basson, C. T. *et al.* Mutations in human *TBX5* cause limb and cardiac malformation in Holt–Oram syndrome. *Nature Genet.* **15**, 30–35 (1997).
- Ahn, D., Ruvinsky, I., Oates, A. C., Silver, L. M. & Ho, R. K. *tbx20*, a new vertebrate *Tbx* gene expressed in the cranial motor neurons and developing cardiovascular structures in zebrafish. *Mech. Dev.* **95**, 253–258 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank A. Oates for helpful suggestions for the construction of the test construct for morpholino oligonucleotides, A. Bruce, A. Oates and I. Skromme for critical comments on the manuscript, I. Ruvinsky for helpful discussions, and H. Dow for fish care. This work was supported by a postdoctoral fellowship from the American Heart Association, Heritage Affiliate (to D.A.), a Cancer Training Grant from NIH (to M.J.K.), and grants from NIH (to L.M.S. and R.K.H.) and NSF (to R.K.H.).

Competing interests statement

The authors declare that they have no competing financial interests

Correspondence and requests for materials should be addressed to D.A. (e-mail: dahn@midway.uchicago.edu).

An abundant erythroid protein that stabilizes free α -haemoglobin

Anthony J. Kihm*, Yi Kong*, Wei Hong*, J. Eric Russell*, Susan Rouda*, Kazuhiko Adachi*, M. Celeste Simon†, Gerd A. Blobel* & Mitchell J. Weiss*

* The Children's Hospital of Philadelphia, Division of Hematology, 34th and Civic Center Blvd, Philadelphia; and the University of Pennsylvania, Philadelphia, Pennsylvania 19104-6160-6160, USA

† Department of Cell and Developmental Biology, Abramson Family Cancer Research Institute, and Howard Hughes Medical Institute, University of Pennsylvania, 421 Curie Boulevard, Philadelphia, Pennsylvania 19104-6160, USA

The development of red blood cells (erythrocytes) is distinguished by high-level production of the oxygen carrier, haemoglobin A (HbA), a heterotetramer of α - and β -haemoglobin subunits. HbA synthesis is coordinated to minimize the accumulation of free subunits that form cytotoxic precipitates^{1–3}. Molecular chaperones that regulate globin subunit stability, folding or assembly have been proposed to exist but have never been identified. Here we identify a protein stabilizing free α -haemoglobin by using a screen for genes induced by the essential erythroid transcription factor GATA-1 (refs 4, 5). Alpha Haemoglobin Stabilizing Protein (AHSP) is an abundant, erythroid-specific protein that forms a stable complex with free α -haemoglobin but not with β -haemoglobin or haemoglobin A ($\alpha_2\beta_2$). Moreover, AHSP specifically protects free α -haemoglobin from precipitation in solution and in live cells. AHSP-gene-ablated mice exhibit reticulocytosis and abnormal erythrocyte morphology with intracellular inclusion bodies that stain positively for denatured haemoglobins. Hence, AHSP is required for normal erythropoiesis, probably acting to block the deleterious effects of free α -haemoglobin precipitation. Accordingly, AHSP gene dosage is predicted to modulate pathological states of α -haemoglobin excess, such as β -thalassaemia.

Many genes and metabolic pathways that participate in haemoglobin synthesis are controlled by GATA-1, a DNA-binding protein that is essential for the survival and maturation of lineage-committed erythroid precursors^{6,7}. Known GATA-1 target genes include those encoding α - and β -globins and haem biosynthetic enzymes⁴. We therefore reasoned that other facets of haemoglobin production and regulation must also be controlled by GATA-1. We performed subtractive hybridization to identify transcripts that were upregulated after genetic rescue of the GATA-1-null erythroid cell line G1E-ER2 (ref. 5). The screen verified known GATA-1 targets and identified numerous other genes of unknown function. Among the latter group, erythroid differentiation-related factor (EDRF) (GenBank accession number AF060220) messenger RNA and protein were strongly and rapidly induced by GATA-1 (Fig. 1a, b). On the basis of the current studies of its cellular function, we propose that EDRF be renamed α -haemoglobin stabilizing protein (AHSP).

Murine AHSP/EDRF RNA encodes a 102-amino-acid protein that is highly conserved in human, pig, cow and rat (not shown) and contains no recognizable signature motifs. AHSP mRNA is expressed exclusively in haematopoietic tissues including bone marrow, spleen and fetal liver (Fig. 1c). RNA *in situ* hybridization analysis of early and mid-gestation murine embryos demonstrated strong expression in embryonic yolk-sac blood islands and fetal liver, respectively (not shown). An analysis of primary haematopoietic colonies by polymerase chain reaction with reverse transcription (RT-PCR) demonstrated that AHSP is expressed in erythroid cells but not in myeloid or megakaryocytic cells (Fig. 1d). As expected for a GATA-1-regulated gene, AHSP mRNA was diminished in GATA-1⁻ erythroid colonies in comparison with wild-type controls. Western blots comparing extracts from erythroid cell lines with known quantities of recombinant AHSP

protein indicated that there are at least 10⁷ AHSP molecules per late erythroid precursor (approximate concentration 0.1 mM; not shown). AHSP is therefore an abundant, erythroid-specific protein that is induced by GATA-1 during erythroid maturation. These results confirm and extend previous studies of AHSP expression⁸.

We searched for clues to AHSP function by identifying interacting partners. Full-length AHSP was used as bait to screen a murine erythroleukaemia (MEL) cell complementary DNA library in a yeast two-hybrid assay. Analysis of 7.4 × 10⁵ colonies yielded 37 positive clones, each encoding α -globin protein. As an independent approach, we demonstrated that glutathione S-transferase (GST)–AHSP fusion protein selectively bound free α -haemoglobin in MEL cell extracts (Fig. 2a).

To determine whether AHSP and haemoglobins interact in solution, we incubated recombinant human AHSP with α - or β -haemoglobins and analysed the reactions by gel-filtration chromatography (Fig. 2b). Eluted haemoglobins were detected by light absorbance at 540 nm, a property conferred by the haem prosthetic group. AHSP was monitored by absorbance at 280 nm. Although pure α -haemoglobin exists primarily as a monomer in solution, incubation with AHSP generated a larger haem-protein species (Fig. 2b) that contained both α -haemoglobin and AHSP by western blotting of collected fractions (not shown). Gel-filtration studies, using various concentrations of AHSP and α -haemoglobin to estimate binding equilibrium, indicated a maximum dissociation constant (K_d) of 0.5 μ M (not shown). Pure β -haemoglobin exists mainly as a tetramer (β_4 , HbH), which did not interact with AHSP in solution (Fig. 2b, bottom two panels).

We studied the specificity of AHSP–haemoglobin interactions by incubating purified haemoglobins with each other or with recombinant AHSP and resolving the complexes by non-denaturing

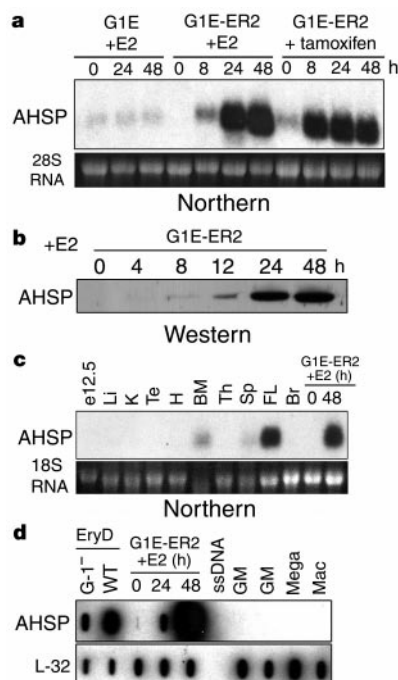


Figure 1 AHSP is a GATA-1-induced erythroid-specific gene. **a**, Northern blot with full-length AHSP complementary DNA as probe. G1E and G1E-ER2 cells were treated with β -oestradiol (E2) or the partial oestrogen receptor agonist tamoxifen. G1E is a GATA-1⁻ erythroid line that undergoes GATA-1-dependent terminal maturation. G1E-ER2 cells are a G1E subclone engineered to express an oestradiol-activated form of GATA-1 (ref. 27). Total RNA per lane, 15 μ g. **b**, Western blot of β -oestradiol-treated G1E-ER2 cells with anti-murine AHSP antibody. Total cellular protein per lane, 10 μ g. **c**, Multi-tissue northern blot with full-length AHSP probe (15 μ g total RNA per lane). e12.5, day 12.5 mouse embryo; Li, adult liver; K, kidney; Te, testis; H, heart; BM, bone marrow; Th, thymus; Sp, spleen; FL, fetal liver; Br, brain. **d**, RT-PCR analysis of AHSP expression in primary haematopoietic colonies. EryD, definitive (adult-type) erythroid colonies derived from wild-type (WT) or GATA-1⁻ (G-1⁻) embryonic stem cells; ssDNA, salmon sperm DNA; GM, granulocyte–macrophage; Mega, megakaryocyte; Mac, macrophage. L-32 is a constitutively expressed ribosomal protein. For comparison, samples were analysed from β -oestradiol (E2)-treated G1E-ER2 cells.

cellulose acetate electrophoresis. Samples analysed in parallel were stained with either benzidine, which detects haem-containing proteins, or Ponceau S, which detects all proteins (Fig. 2c, d). Free α -haemoglobin migrated slowly and stained with both reagents (Fig. 2c, lane 1), whereas AHSP migrated as a fast-moving doublet

that stained only with Ponceau S (Fig. 2d, lane 7). Incubation of AHSP with α -haemoglobin resulted in the formation of an α -haemoglobin-AHSP complex (Fig. 2c, lanes 2–5) that stained with benzidine and contained AHSP, as determined by western blotting of protein extracted from the cellulose acetate matrix. AHSP did not

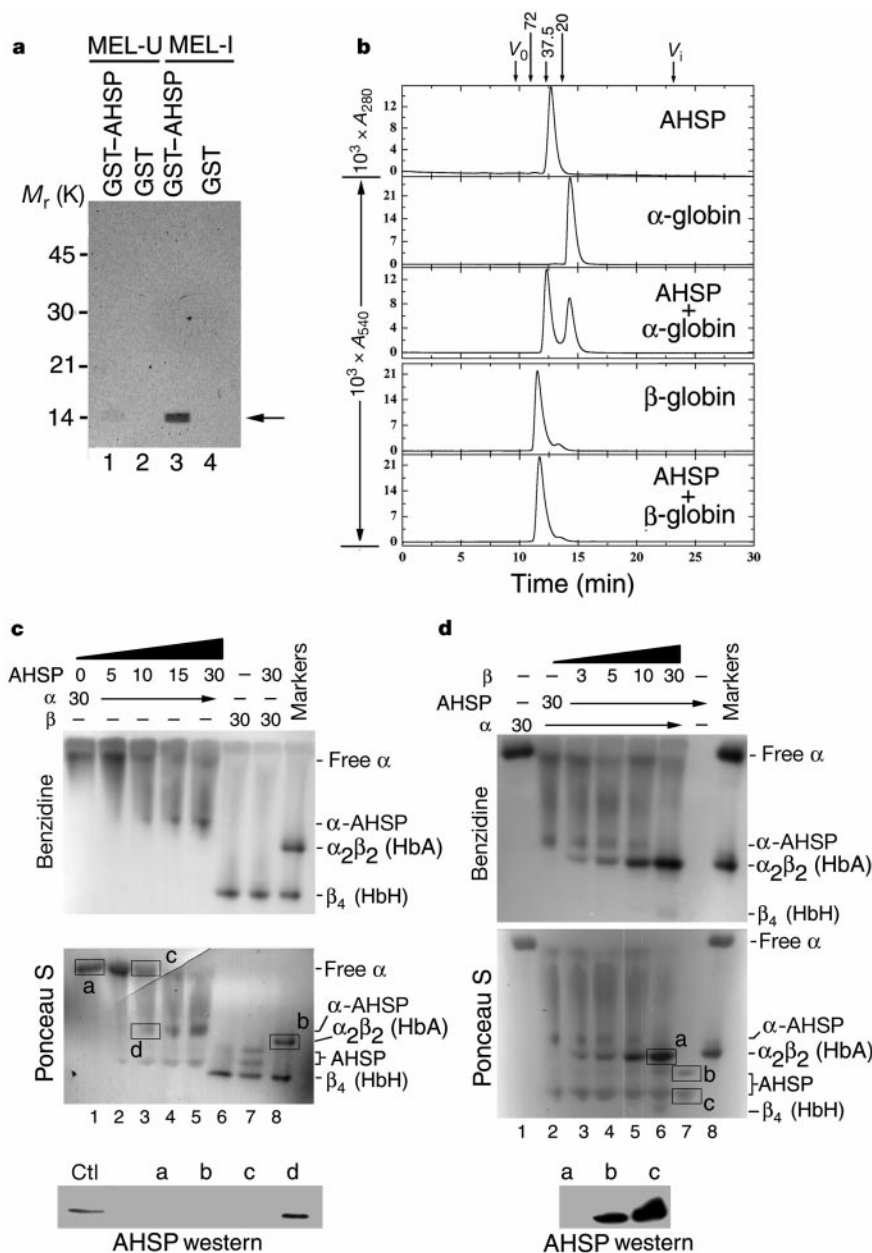


Figure 2 AHSP binds free α -haemoglobin but not β -haemoglobin. **a**, GST-AHSP pull-down analysis. AHSP-GST or GST crosslinked to agarose beads were incubated with extracts from MEL cells. Bound protein was resolved on a 12.5% SDS-polyacrylamide gel and detected with Coomassie blue. U, uninduced; I, DMSO-induced for 72 h. The prominent bound protein (arrow) was determined as being α -haemoglobin by Edman sequence analysis. **b**, Size-exclusion chromatography. Mixtures containing equimolar ratios of recombinant human AHSP and purified human carboxy (CO)- α - or β -haemoglobins were analysed by gel filtration. AHSP (top panel) was detected by light absorbance at 280 nm; haemoglobins were detected by absorbance at 540 nm. Arrows indicate molecular size standards (shown as $10^{-3} \times$ molecular masses). In solution, purified α -haemoglobin exists mainly as a monomer ($M_r \sim 16$ K); β -haemoglobin exists mainly as a tetramer (HbH, $M_r \sim 64$ K). The retention time observed for AHSP predicts a M_r of ~ 35 K, significantly greater than the 12K predicted by the primary amino acid

sequence. This anomaly could be a consequence of either multimerization or an asymmetric shape. **c**, Cellulose acetate electrophoresis. AHSP was incubated with purified CO- α - or CO- β -haemoglobin and analysed. Duplicate strips were stained with benzidine to identify haem-containing proteins, or with Ponceau S to stain all proteins. Concentrations of AHSP and haemoglobins (μ M) are indicated at the top. Bands a–d were eluted from the Ponceau-S-stained cellulose acetate strip and analysed by western blotting with anti-AHSP antibody (bottom panel). Extract from oestradiol-treated G1E-ER2 cells (Ctl) served as a positive control for AHSP protein. **d**, AHSP and β -haemoglobin compete for α -haemoglobin binding. AHSP and CO- α -haemoglobin were preincubated for 1 h, then CO- β -haemoglobin was added at the indicated concentrations (μ M) and the samples were incubated for an additional 30 min. Cellulose acetate electrophoresis and western blotting were performed as described in **c**.

affect the mobility of haemoglobin H (Fig. 2c, lane 7) or haemoglobin A (not shown). Taken together, our gel-filtration and cellulose acetate electrophoretic studies demonstrate that AHSP forms a specific and stable complex with α -haemoglobin.

The α -haemoglobin–AHSP complex was disrupted by the addition of β -haemoglobin, with the concomitant appearance of haemoglobin A ($\alpha_2\beta_2$), which did not contain AHSP as detected by western blotting (Fig. 2d, lanes 3–6). Hence, AHSP and β -haemoglobin compete for α -haemoglobin binding, with β -haemoglobin binding more tightly.

Next, we investigated whether AHSP alters functionally important physico-chemical properties of α -haemoglobin. Free α -haemoglobin is highly unstable in cells, forming inclusion bodies that are detected by staining with crystal violet (Heinz bodies). These inclusion bodies decrease erythrocyte lifespan by damaging and destabilizing the cell membrane^{2,9–12}. Precipitated α -haemoglobin can also cause apoptosis of erythroid precursors in the bone marrow and spleen^{3,13–15}. These events underlie the pathophysiology of ineffective erythropoiesis in β -thalassaemia, in which defects in β -haemoglobin production lead to an excess of α -subunits and the formation of inclusion bodies. The precipitation of α -haemoglobin *in vivo* is triggered by oxidation of haem-bound iron, initiating a series of structural perturbations that culminate in protein denaturation. This molecular cascade is modelled *in vitro* by exposing free α -haemoglobin to the oxidant ferricyanide, which induces its rapid precipitation (Fig. 3a)¹⁶. Remarkably, AHSP, but not the control proteins IgG or GST, blocked oxidant-induced α -haemoglobin precipitation in a dose-dependent fashion (Fig. 3a, b, and not shown). In contrast, oxidant-induced precipitation of β -haemoglobin (Fig. 3c) was not inhibited by AHSP, indicating that its stabilizing effects are specific for haemoglobin subunits. Spectrophotometric analysis demonstrated that AHSP-bound α -haemoglobin was oxidized by ferricyanide (Fig. 3d). AHSP therefore prevents the precipitation of oxidized α -haemoglobin rather than

simply acting as an anti-oxidant.

To determine whether AHSP and α -haemoglobin interact *in vivo*, we expressed epitope-tagged versions of each protein in COS cells (Fig. 4). Coexpressed AHSP and α -haemoglobin immunoprecipitated together from cell extracts, which is consistent with complex formation *in vivo* (Fig. 4a). We then detected protein distribution by indirect immunofluorescence with the use of laser confocal microscopy (Fig. 4b). When expressed alone, α -haemoglobin precipitated in a punctate pattern, in a similar manner to inclusion body formation in β -thalassaemic erythroid precursors that contain excess free α -haemoglobin¹⁷. In contrast, α -haemoglobin was distributed homogeneously throughout the cytoplasm when expressed with AHSP, showing that AHSP can prevent α -haemoglobin precipitation in intact cells.

To investigate the requirement for AHSP *in vivo*, we deleted the entire protein-coding region of the gene in mice (Supplementary Information, Fig. S1). AHSP^{-/-} mice were born at expected mendelian ratios and displayed grossly normal growth and development (not shown). Although the mice exhibited normal blood haemoglobin concentrations and haematocrits (Fig. 5a, and not shown), their reticulocyte counts were elevated about threefold (Fig. 5b), indicating a shortened erythrocyte half-life. AHSP^{-/-} erythrocytes exhibited an abnormal spiculated morphology (Fig. 5c, top right panel) that has also been observed in erythroid cells of β -thalassaemic mice^{18,19}. Moreover, AHSP^{-/-} erythrocytes contained denatured haemoglobin inclusions (Heinz bodies) that stained with crystal violet (Fig. 5c, bottom right panel). Thus, the loss of AHSP leads to defective haemoglobin metabolism *in vivo*. Considering our biochemical data, the erythrocyte abnormalities observed in AHSP^{-/-} mice probably result from a pathological accumulation of precipitated α -haemoglobin. In this regard, erythroid precursors contain a small pool of excess free α -haemoglobin^{20,21} whose stability and toxicity might be limited by direct interaction with AHSP. Thus, AHSP could interact transiently with free α -haemo-

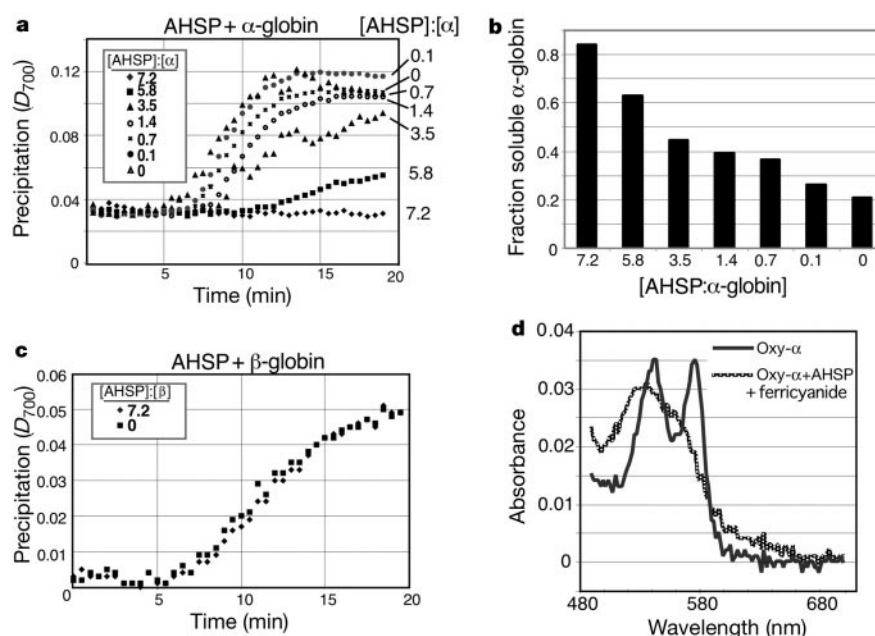


Figure 3 AHSP prevents oxidant-induced precipitation of α -haemoglobin in solution. **a**, Human AHSP was preincubated with purified oxygenated (oxy) α -haemoglobin at the indicated molar ratios for 60 min. Potassium ferricyanide (50 μ M final concentration) was added at zero time to induce haem oxidation. Protein precipitation was monitored by light scattering at 700 nm. **b**, α -Haemoglobin reactions from **a** were centrifuged 60 min after the addition of ferricyanide. Soluble α -haemoglobin remaining in the supernatant was

quantified by light absorbance at 540 nm. The y-axis shows the fraction of the original α -haemoglobin remaining in solution. **c**, Oxy- β -haemoglobin was preincubated with AHSP, oxidized by ferricyanide and analysed as described in **a**. **d**, Absorption spectrum of oxygenated oxy- α -haemoglobin compared with oxy- α -haemoglobin preincubated with a 7.2-fold excess of AHSP, then treated with ferricyanide. The latter sample shows spectrophotometric changes consistent with oxidation of the haem-associated iron²⁹.

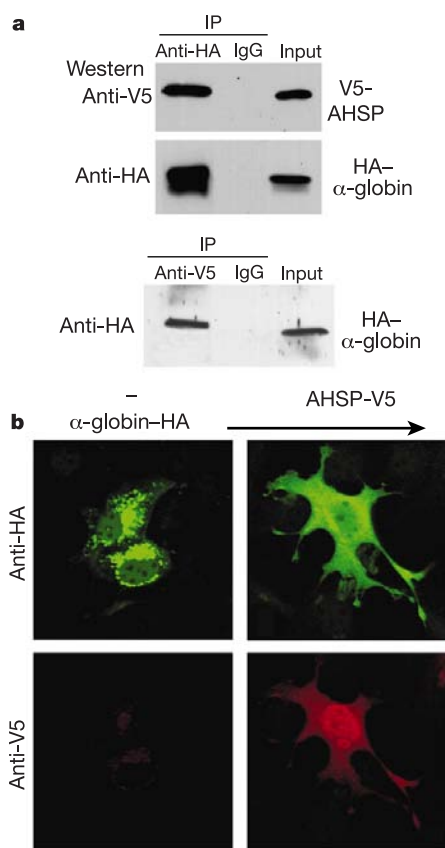


Figure 4 AHSP- α -haemoglobin interactions in mammalian cells. **a**, Immunoprecipitation of AHSP with α -haemoglobin expressed in COS cells. V5-epitope-tagged murine AHSP and haemagglutinin (HA)-tagged murine α -haemoglobin were expressed in COS cells and extracts were analysed by immunoprecipitation (IP) followed by western blot analysis. **b**, AHSP prevents α -haemoglobin precipitation *in vivo*. COS cells expressing HA- α -haemoglobin alone or in combination with V5-AHSP, were stained with antibodies against V5 and HA and analysed by indirect immunofluorescence. Original magnification $\times 400$.

globin to maintain it in a quasi-native conformation before association with β -haemoglobin. The reticulocyte counts in AHSP^{+/-} mice were mildly elevated (Fig. 5b), indicating that AHSP haploinsufficiency might produce a subtle erythroid phenotype. Further studies are required to address this possibility.

It has recently been shown that AHSP (EDRF) mRNA is specifically downregulated in spleen, bone marrow and blood of animals with transmissible spongiform encephalopathies (TSEs, prion disease)⁸. Although haematopoietic tissues are believed to participate in TSE pathogenesis, the role of AHSP in this process is unknown. TSEs are caused by refolding of the cellular protein PrP^C into the disease-associated form, PrP^{Sc}, which forms toxic aggregates in the central nervous system²². In principle, this structural rearrangement could be influenced directly or indirectly by AHSP through its demonstrated ability to alter protein aggregation.

Most cells use two general strategies for neutralizing unstable proteins: proteolysis and stabilization by molecular chaperones²³. Proteolytic pathways exist that degrade free α -haemoglobin in erythroid precursors^{24,25}. Here we provide direct evidence for an erythroid-specific molecular chaperone for free α -haemoglobin. These two complementary mechanisms, acting independently or coordinately, could explain how individuals with the β -thalassaemia trait tolerate a deficit in β -haemoglobin production with no α -haemoglobin precipitation in erythroid precursors²⁶. Our model predicts that full or partial loss of AHSP function would exacerbate mild or intermediate forms of β -thalassaemia. Conversely, delivery

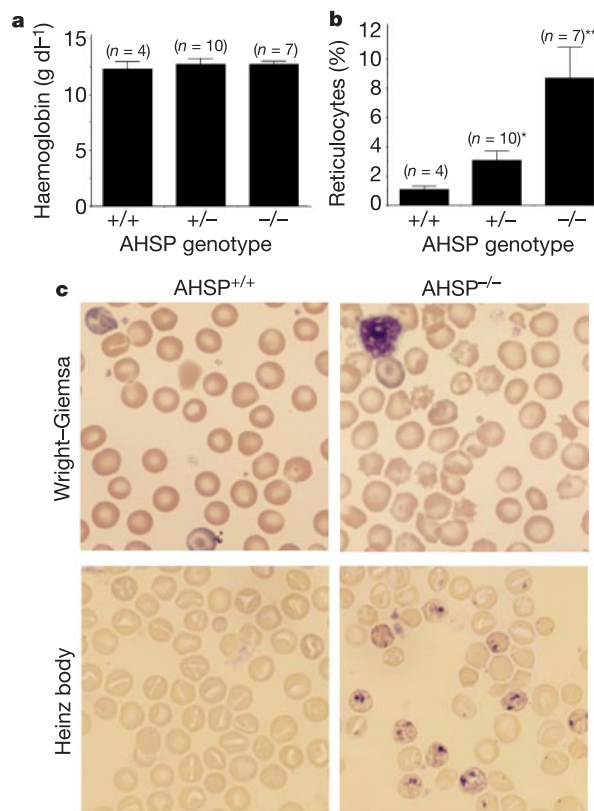


Figure 5 Erythrocyte abnormalities in AHSP-gene-targeted mice. **a**, Blood haemoglobin concentrations expressed as means $\pm$ s.e.m. for mutant and wild-type littermates. No significant differences were observed in other erythrocyte indices including haematocrit, erythrocyte number, mean corpuscular volume, mean corpuscular haemoglobin and mean corpuscular haemoglobin concentration (not shown). **b**, Blood reticulocyte counts expressed as percentages of total erythrocytes. Bars indicate means $\pm$ s.e.m. Single asterisk indicates $P < 0.05$, double asterisk indicates $P < 0.01$. **c**, Wright-Giemsa stain (upper panels) shows abnormal, spiculated erythrocytes in AHSP^{-/-} mice; crystal violet stain (lower panels) reveals Heinz-body inclusions in AHSP^{-/-} erythrocytes.

of AHSP or functional analogues are predicted to decrease ineffective erythropoiesis caused by the precipitation of α -haemoglobin in β -thalassaemia. □

Methods

Cell lines

G1E-ER2 cells were cultured and induced to differentiate with β -oestradiol as described²⁷. MEL cells were grown in Dulbecco's modified Eagle's medium (Gibco-BRL) supplemented with 10% fetal bovine serum. To induce erythroid maturation, cells were seeded at a density of 10^5 ml⁻¹ in medium containing 2% dimethyl sulphoxide (DMSO).

Plasmids

Human and mouse AHSP cDNAs were isolated by RT-PCR from fetal liver with SuperScriptII reverse transcriptase (Gibco-BRL) and *Pfu* polymerase (Stratagene). PCR primers were as follows: mouse AHSP, 5'-GGATGAGGCGGGCTCAGCACCATTAT-3' and 5'-TCGGCTCACAAACCCCAAGATTC-3'; human AHSP, 5'-TGAAGGCAGATGGCTCTTC-3' and 5'-GTCCTGAGCTAGGAGGAGG-3'. PCR conditions were 94 °C for 30 s, 55 °C for 30 s, 72 °C for 45 s (30 cycles). cDNAs derived by RT-PCR were sequenced on both strands.

Proteins

GST-fused mouse and human AHSP were prepared with the vector pGEX 6P-1 (Amersham-Pharmacia) in accordance with the manufacturer's instructions. Human haemoglobins were prepared as described²⁸. Carboxyhaemoglobins were used for protein interaction studies (Fig. 2b-d) because this enhances protein stability by limiting the generation of oxidative products. Oxy- α - and oxy- β -haemoglobins were used for studies on protein stability (Fig. 3).

Additional methods for protein, nucleic acid and animal studies are described in Supplementary Information.

Received 11 February; accepted 18 March 2002; doi:10.1038/nature00803.

1. Baglioni, C. Chromosomal and cytoplasmic regulation of haemoglobin synthesis. *Bibl. Haemat.* **29**, 1056–1063 (1966).
2. Nathan, D. G. & Gunn, R. B. Thalassemia: the consequences of unbalanced hemoglobin synthesis. *Am. J. Med.* **41**, 815–830 (1966).
3. Rachmilewitz, E. & Schrier, S. L. in *Disorders of Haemoglobin* (eds Steinberg, M. H., Forget, B. G., Higgs, D. R. & Nagel, R. L.) 233–251 (Cambridge Univ. Press, Cambridge, 2001).
4. Weiss, M. J. & Orkin, S. H. GATA transcription factors: Key regulators of hematopoiesis. *Exp. Hematol.* **23**, 99–107 (1995).
5. Shirihai, O. S., Gregory, T., Yu, C., Orkin, S. H. & Weiss, M. J. ABC-me, a novel mitochondrial transporter induced by GATA-1 during erythroid differentiation. *EMBO J.* **19**, 2492–2502 (2000).
6. Weiss, M. J., Keller, G. & Orkin, S. H. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1⁺ embryonic stem cells. *Genes Dev.* **8**, 1184–1197 (1994).
7. Fujiwara, Y., Browne, C. P., Cunneiff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl Acad. Sci. USA* **93**, 12355–12358 (1996).
8. Miele, G., Manson, J. & Clinton, M. A novel erythroid-specific marker of transmissible spongiform encephalopathies. *Nature Med.* **7**, 361–364 (2001).
9. Nathan, D. G., Stossel, T. B., Gunn, R. B., Zarkowsky, H. S. & Laforet, M. T. Influence of hemoglobin precipitation on erythrocyte metabolism in alpha and beta thalassemia. *J. Clin. Invest.* **48**, 33–41 (1969).
10. Shinar, E., Shalev, O., Rachmilewitz, E. A. & Schrier, S. L. Erythrocyte membrane skeleton abnormalities in severe beta-thalassemia. *Blood* **70**, 158–164 (1987).
11. Sorensen, S., Rubin, E., Polster, H., Mohandas, N. & Schrier, S. The role of membrane skeletal-associated alpha-globin in the pathophysiology of beta-thalassemia. *Blood* **75**, 1333–1336 (1990).
12. Yuan, J. *et al.* The instability of the membrane skeleton in thalassemic red blood cells. *Blood* **86**, 3945–3950 (1995).
13. Yuan, J. *et al.* Accelerated programmed cell death (apoptosis) in erythroid precursors of patients with severe beta-thalassemia (Cooley's anemia). *Blood* **82**, 374–377 (1993).
14. Pootrakul, P. *et al.* A correlation of erythrokinetics, ineffective erythropoiesis, and erythroid precursor apoptosis in Thai patients with thalassemia. *Blood* **96**, 2606–2612 (2000).
15. Centis, F. *et al.* The importance of erythroid expansion in determining the extent of apoptosis in erythroid precursors in patients with beta-thalassemia major. *Blood* **96**, 3624–3629 (2000).
16. Rachmilewitz, E. A., Peisach, J. & Blumberg, W. E. Studies on the stability of oxyhemoglobin A and its constituent chains and their derivatives. *J. Biol. Chem.* **246**, 3356–3366 (1971).
17. Wickramasinghe, S. N. & Hughes, M. Ultrastructural studies of erythropoiesis in beta-thalassemia trait. *Br. J. Haematol.* **46**, 401–407 (1980).
18. Ciavatta, D. J., Ryan, T. M., Farmer, S. C. & Townes, T. M. Mouse model of human beta zero thalassemia: targeted deletion of the mouse beta maj- and beta min-globin genes in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **92**, 9259–9263 (1995).
19. Yang, B. *et al.* A mouse model for beta (0)-thalassemia. *Proc. Natl Acad. Sci. USA* **92**, 11608–11612 (1995).
20. Shaeffer, J. R. Evidence for soluble alpha-chains as intermediates in hemoglobin synthesis in the rabbit reticulocyte. *Biochem. Biophys. Res. Commun.* **28**, 647–652 (1967).
21. Tavill, A. S., Grayzel, A. L., Vanderhoff, G. A. & London, I. M. The control of hemoglobin synthesis. *Trans. Assoc. Am. Physic.* **80**, 305–313 (1967).
22. Prusiner, S. B., Scott, M. R., DeArmond, S. J. & Cohen, F. E. Prion protein biology. *Cell* **93**, 337–348 (1998).
23. Wickner, S., Maurizi, M. R. & Gottesman, S. Posttranslational quality control: folding, refolding, and degrading proteins. *Science* **286**, 1888–1893 (1999).
24. Bank, A. & O'Donnell, J. V. Intracellular loss of free α chains in β -thalassaemia. *Nature* **222**, 295–296 (1969).
25. Shaeffer, J. R. Turnover of excess hemoglobin α chains in β -thalassemic cells is ATP-dependent. *J. Biol. Chem.* **258**, 13172–13177 (1983).
26. Huang, S.-C. & Benz, E. J. in *Disorders of Haemoglobin* (eds Steinberg, M. H., Forget, B. G., Higgs, D. R. & Nagel, R. L.) 146–173 (Cambridge Univ. Press, Cambridge, 2001).
27. Gregory, T. *et al.* GATA-1 and erythropoietin cooperate to promote erythroid cell survival by regulating bcl-x_L expression. *Blood* **94**, 87–96 (1999).
28. Ikeda-Saito, M., Inubushi, T. & Yonetani, T. in *Hemoglobins* (eds Antonini, E., Rossi-Bernardi, L. & Chiancone, E.) 113–121 (Academic, New York, 1981).
29. Rachmilewitz, E. A. Denaturation of the normal and abnormal hemoglobin molecule. *Semin. Hematol.* **11**, 441–462 (1974).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank S. Krishnaswamy for advice and assistance in biochemical studies and use of equipment, J. Yu and Z. He for technical assistance, D. Speicher for assistance with protein sequencing, C. Clendenin for assistance with mouse blastocyst injections, T. Labosky for providing embryonic stem cells, and D. Nathan, D. Pellman and W. Englander for helpful discussions and review of the manuscript. This work was funded by grants from the Cooley's Anemia Foundation (M.J.W.), the Unico Foundation (M.J.W.) and the National Institutes of Health (M.J.W., J.E.R. and G.A.B.). M.J.W. is a recipient of The American Society of Hematology Junior Faculty Award.

Competing interests statement

The authors declare that they have no competing financial interests

Correspondence and requests for materials should be addressed to M.J.W. (e-mail: weissmi@email.chop.edu).

Regulation of *Arabidopsis* cryptochrome 2 by blue-light-dependent phosphorylation

Dror Shalitin*, Hongyun Yang*, Todd C. Mockler*, Maskit Maymon*, Hongwei Guo†, Garry C. Whitelam‡ & Chentao Lin*

* Department of Molecular, Cell & Developmental Biology, University of California, Los Angeles, California 90095, USA

‡ Biology Department, University of Leicester, Leicester LE1 7RH, UK

† Present address: Genomic Analysis Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, California 92037, USA.

Cryptochromes are blue/ultraviolet-A light receptors that mediate various light responses in plants and animals^{1,2}. But the initial photochemical reaction of cryptochrome is still unclear. For example, although most photoreceptors are known to undergo light-dependent protein modification such as phosphorylation^{3,4}, no blue-light dependent phosphorylation has been reported for a cryptochrome. *Arabidopsis* cryptochrome 2 (cry2) mediates light regulation of seedling development and photoperiodic flowering^{5,6}. The physiological activity and cellular level of cry2 protein are light-dependent^{5–8}, and protein–protein interactions are important for cry2 function^{9,10}. Here we report that cry2 undergoes a blue-light-dependent phosphorylation, and that cry2 phosphorylation is associated with its

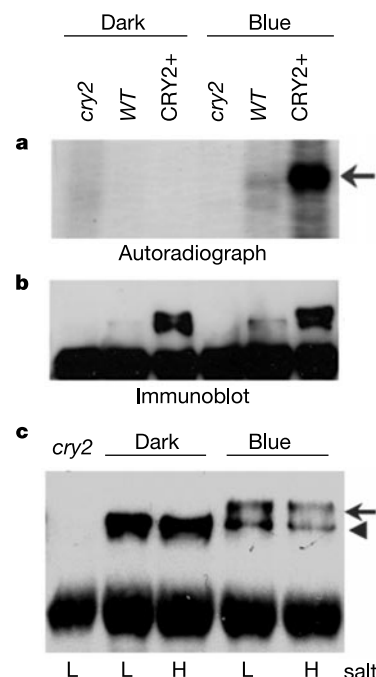


Figure 1 Blue-light-dependent phosphorylation of *Arabidopsis* cry2. **a**, Six-day-old seedlings were incubated with radioactive ³²P, exposed to blue light (5 μ mol m⁻² s⁻¹) for 10 min (Blue), or kept in the dark (Dark). Immunoprecipitation obtained using anti-CRY2 antibodies were separated in a 10% Laemmli gel, blotted, and exposed to an X-ray film. **b**, An immunoblot of **a** was probed with anti-CRY2 antibodies. **c**, Samples of 5-day-old wild-type seedlings with (Blue) or without (Dark) a 15-min blue-light treatment (5 μ mol m⁻² s⁻¹) were immunoprecipitated with anti-CRY2 antibodies in low-salt (100 mM NaCl, L) or high-salt buffer (200 mM NaCl, H) fractionated in a 10% NuPAGE gel, blotted, and probed with anti-CRY2 antibodies. In Figs 1–4, arrows indicate phosphorylated cry2; arrowheads indicate unphosphorylated cry2.

function and regulation. Our results suggest that, in the absence of light, cry2 remains unphosphorylated, inactive and stable; absorption of blue light induces the phosphorylation of cry2, triggering photomorphogenic responses and eventually degradation of the photoreceptor.

To test a hypothesis that the absorption of photons results in a biochemical modification of a cryptochrome molecule, we investigated whether *Arabidopsis* cry2 underwent a blue-light-induced phosphorylation *in vivo*. We fed etiolated (dark-grown) seedlings with radioactive phosphate ($^{32}\text{PO}_4^-$) before they were exposed to blue light, and then examined the cry2 phosphorylation. Autoradiography and immunoblot analysis of proteins purified by immunoprecipitation using anti-CRY2 antibody demonstrated that cry2 is phosphorylated in response to blue light (Fig. 1a, b). No radioactive cry2 was found in seedlings incubated in ^{32}P for 2 h in the dark, but ^{32}P -labelled cry2 was detected within 15 min of blue-light treatment. The level of radioactively labelled cry2 correlated with the amount of cry2 protein found in different genotypes (Fig. 1a, b). No cry2 or radioactively labelled immunoprecipitation product was detected in the cry2 mutant, but both were found at higher levels in transgenic plants overexpressing cry2 than in the wild type. Although a similar level of cry2 is detected by immunoblot in CRY2+ transgenic plants irrespective of a brief light treatment, the ^{32}P -labelled cry2 was detected only in blue-light-treated seedlings (Fig. 1a, b).

Protein phosphorylation often causes slow migration of the phosphorylated protein in a SDS-polyacrylamide gel electrophor-

esis (PAGE) gel, allowing a more convenient analysis than the *in vivo* labelling experiments. A migration shift was only occasionally observed for the cry2 protein analysed using conventional Laemmli SDS-PAGE gels¹¹ (data not shown). Using a commercially available NuPAGE gel system (see Methods), two CRY2 isoforms were more consistently found in seedlings exposed to blue light, a fast-migrating band and a slow-migrating band, whereas only the fast-migrating CRY2 band was found in etiolated seedlings (Fig. 1c). The presence of the slow-migrating band found in blue-light-treated seedlings was not affected by a change in the ionic strength of the buffer used in the immunoprecipitation reaction (Fig. 1c), but it was eliminated by treating the sample with lambda phosphatase, which nonspecifically dephosphorylates proteins (see below).

We concluded that the slow-migrating band found in seedlings exposed to blue light is phosphorylated cry2, whereas the fast-migrating band from dark-grown or light-treated seedlings is unphosphorylated cry2. A significant fraction of cry2 remained unphosphorylated in plants exposed to blue light (Figs 1c and Fig. 2), suggesting that light-grown plants contain both phosphorylated and unphosphorylated cry2. Provided that one isoform is physiologically more active than the other one, the presence of both isoforms of cry2 in the cell may allow the total activity of cry2 to be adjusted rapidly in response to changing light conditions.

The light-induced phosphorylation of cry2 was detected shortly after seedlings were exposed to blue light (Fig. 2a, b). At a given fluence rate of light, the relative level of phosphorylated cry2 increased initially with the exposure time, but it decreased after a

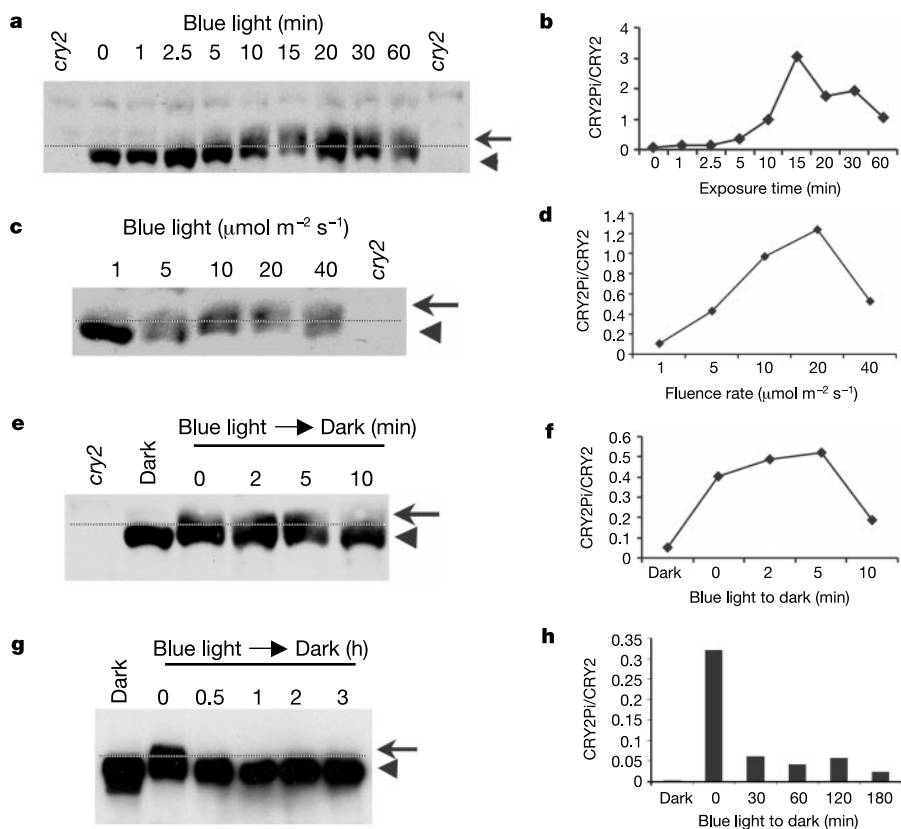


Figure 2 Kinetics analyses of light-dependent cry2 phosphorylation. Immunoblots in **a**, **c**, **e** and **g** were prepared in 10% NUPAGE gels and probed with anti-CRY2 antibodies. Samples of wild-type (**a**, **c**, **e**) or transgenic plants overexpressing CRY2 (**g**) were analysed. Six-day-old dark-grown seedlings were treated in the following ways: **a**, **b**, exposed to blue light ($5 \mu\text{mol m}^{-2} \text{s}^{-1}$) for the periods indicated; **c**, **d**, exposed to blue light for 15 min at the fluence rates indicated; **e**, **f** exposed to blue light ($5 \mu\text{mol m}^{-2} \text{s}^{-1}$)

for 10 min, and then transferred to the dark for the time indicated; **g**, **h**, exposed to blue light ($5 \mu\text{mol m}^{-2} \text{s}^{-1}$) for 20 min, and then transferred to the dark for the time indicated. The relative band intensities of phosphorylated cry2 (above the line) versus that of unphosphorylated cry2 (below) were presented as CRY2Pi/CRY2, and plotted in **b**, **d**, **f** and **h**, for the corresponding immunoblots shown in **a**, **c**, **e** and **g**, respectively.

prolonged blue-light treatment. The cry2 phosphorylation is fluence-rate-dependent (Fig. 2c, d). For a given time period of blue light treatment, relatively more cry2 phosphorylation was detected at higher fluence rates of blue light. However, a further increase of fluence rate resulted in a decrease in the relative amount of phosphorylated cry2. The level of phosphorylated cry2 decreased immediately after wild-type seedlings were transferred from blue light to dark, with little phosphorylated cry2 left shortly after plants were transferred to dark (Fig. 2e, f). Similarly, when CRY2 + transgenic seedlings were examined, the phosphorylated cry2 disappeared after blue-light-treated plants were transferred to dark (Fig. 2g, h).

The kinetic features of cry2 phosphorylation suggest that the phosphorylated cry2 is degraded, because the level of both total cry2 and phosphorylated cry2 declined in response to either increased exposure time or increased light intensity (Fig. 2)^{5,7,12}. To further investigate this hypothesis, we screened for an *Arabidopsis* mutant that is impaired in cry2 degradation. Because a defect in the degradation of a photoreceptor might result in increased light responses, we examined blue-light-dependent cry2 degradation in *Arabidopsis* mutants that are known to exhibit either constitutive or hypersensitive light responses, including *cop1*, *cop9*, *det1*, *det2*, *det3*, *fus6* (*cop11*), *fus9* (*cop10*), *eid1*, and *sub1*^{13–16}. An apparent defect of cry2 degradation was found only in the *cop1* mutant (Fig. 3). The *cop1* mutants are constitutively photomorphogenic¹⁷. Most *cop1* mutants are lethal except the two weak alleles, *cop1-4* and *cop1-6*^{17,18}. Figure 3a showed that the cry2 degradation is partially impaired in both *cop1* mutants tested; higher levels of cry2 protein were detected in *cop1-4* and *cop1-6* seedlings than in the wild type under different blue-light conditions (Fig. 3a). In contrast, no obvious abnormality

in light-induced cry2 degradation was found in the *det1-1* mutant (Fig. 3a), which, like the *cop1* mutant, also exhibits a light-independent de-etiolation response^{17,19}.

The *COP1* gene encodes a putative subunit of an E3 ubiquitin ligase complex associated with proteolysis of the light-signalling transcription factor HY5 (ref. 20), and COP1 has been shown to physically interact with cryptochromes^{10,21}. The exact role of COP1 in cry2 degradation is not clear, because blue-light-dependent cry2 degradation was still observed in *cop1* null alleles (data not shown). Nevertheless, regardless of the exact role of COP1 in cry2 degradation, the impairment of cry2 degradation in the *cop1* weak alleles allowed us to explore further the relationship between cry2 phosphorylation and cry2 degradation. We compared cry2 phosphorylation and cry2 degradation in wild-type plants and *cop1-6* mutant plants exposed to blue light for various time periods (Fig. 3b–d). It was found that a similar amount of unphosphorylated cry2 was detected in dark-grown *cop1-6* or wild-type seedlings. However, in seedlings exposed to blue light, the relative level of phosphorylated cry2 was significantly higher in the *cop1-6* mutant than in the wild type (Fig. 3b). A quantitative comparison of the relative amount of phosphorylated cry2 and the total amount of cry2 in the *cop1-6* mutant showed a correlation of a decrease of cry2 degradation and an increase of the relative abundance of phosphorylated cry2 (Fig. 3c, d).

Phytochromes have been shown to have protein kinase activity and to interact physically with cryptochromes^{9,22,23}. Moreover, the function of *Arabidopsis* cry2 in the regulation of photoperiodic flowering is dependent on phyB²⁴, and cry2 has been shown to directly interact with phyB *in vivo*⁹. To investigate a possible involvement of phytochromes in cry2 phosphorylation, we first tested whether red light might induce cry2 phosphorylation. No cry2 phosphorylation was detected in seedlings exposed to red light (Fig. 4a) or far-red light (data not shown), which demonstrates the wavelength specificity of cry2 phosphorylation. We then investigated whether phytochromes might catalyse blue-light-dependent cry2 phosphorylation by testing cry2 phosphorylation in various phytochrome mutant lines. None of the monogenic phytochrome mutants tested, including *phyA*, *phyB*, *phyD*, *phyE* or the *cry1* mutant, had apparent defects in cry2 phosphorylation (Fig. 4c, and data not shown). To test a possibility that different phytochromes may act as cry2 kinases in a functionally redundant manner, we examined cry2 phosphorylation in the *phyAB* double mutant, and the *phyABD* and *phyBDE* triple mutants^{25,26} (G.C.W., unpublished work), and the *hy1* mutant, which is impaired in phytochrome chromophore biosynthesis²⁷. Again, cry2 phosphorylation was detected in all of these phytochrome mutants (Fig. 4b, c). A protein phosphatase treatment confirmed that the slow-migrating band of cry2 was indeed the phosphorylated form of cry2 (Fig. 4c).

It is somewhat surprising that the *phyBDE* triple mutant showed no obvious defects in cry2 phosphorylation, because phyB, phyD and phyE share a partial functional redundancy^{25,26} and phyB can directly interact with cry2 (ref. 9). It seems unlikely that any single phytochrome species tested here is solely responsible for cry2 phosphorylation. However, our results do not exclude a possible phytochrome involvement in cryptochrome phosphorylation, because not all phytochrome mutants, and in particular, a null mutant lacking all five phytochromes, are currently available. Alternatively, cry2 phosphorylation may be catalysed by a non-phytochrome protein kinase.

The blue-light dependency of cry2 phosphorylation and a bell-shaped curve of the fluence-rate response of cry2 phosphorylation (Fig. 2d) indicate that the phosphorylated form of cry2 may represent physiologically active cry2. This is because the activity of cry2 in mediating blue-light-specific de-etiolation responses is limited primarily to low light-intensity conditions⁵, in which relatively more phosphorylated cry2 was observed (Fig. 2a–d). It

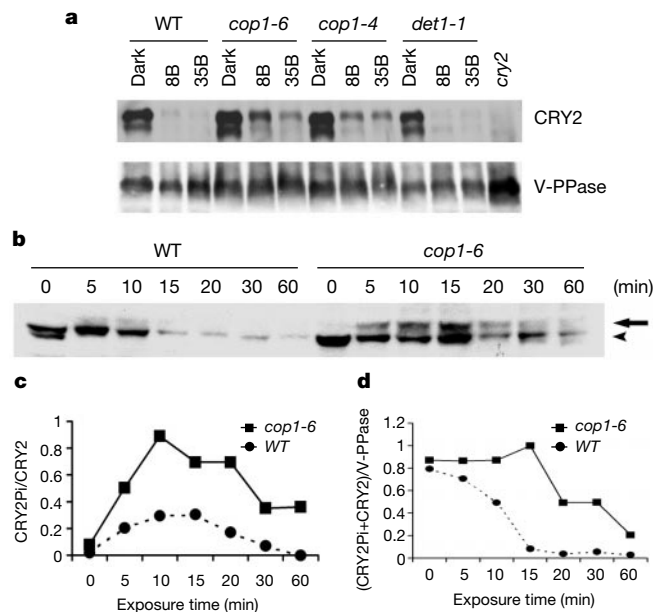


Figure 3 Accumulation of phosphorylated cry2 and impairment of cry2 turnover in *cop1* mutants in response to blue light. **a**, Seven-day-old seedlings were exposed to blue light ($8 \mu\text{mol m}^{-2} \text{s}^{-1}$ (8B) or $35 \mu\text{mol m}^{-2} \text{s}^{-1}$ (35B)) for two hours, and protein extracts were fractionated in a 10% Laemmli gel. The immunoblot was probed with anti-CRY2 antibody (CRY2), stripped and re-probed with antibodies against vacuole pyrophosphatase (V-PPase) for the loading control. **b**, Samples of 7-day-old wild-type (WT) or *cop1-6* seedlings, exposed to blue light ($5 \mu\text{mol m}^{-2} \text{s}^{-1}$) for the time periods indicated, were fractionated in a 10% NuPAGE gel and the immunoblot probed with anti-CRY2. **c**, Plots showing relative levels of phosphorylated cry2 versus unphosphorylated cry2 (CRY2Pi/CRY2) of samples from **b**. **d**, Plots showing total cry2 in samples from **b**, normalized using V-PPase signals (not shown) in the respective lane ((CRY2Pi+CRY2)/V-PPase).

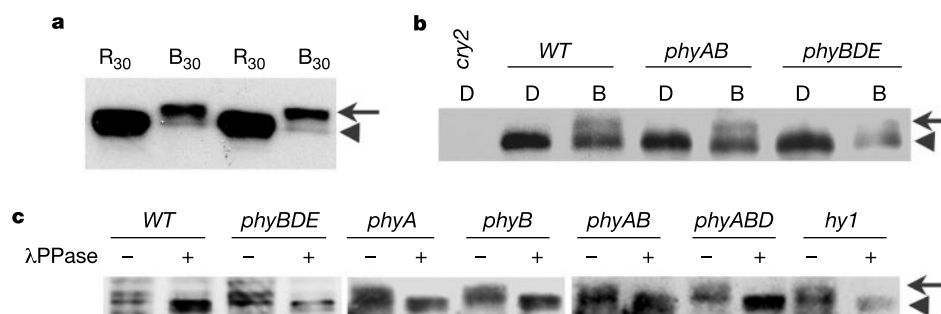


Figure 4 The lack of effect of red light or phytochrome mutations on *cry2* phosphorylation. **a**, CRY2 overexpressing transgenic plants were exposed to red light ($50 \mu\text{mol m}^{-2} \text{s}^{-1}$), or blue light ($5 \mu\text{mol m}^{-2} \text{s}^{-1}$), for 30 min (R_{30} , B_{30}) or 60 min (R_{60} , B_{60}). **b**, Seedlings of the indicated genotypes were exposed to blue light ($5 \mu\text{mol m}^{-2} \text{s}^{-1}$) for 10 min (B) or left

in dark (D). **c**, Seedlings of the indicated genotypes were exposed to blue light ($5 \mu\text{mol m}^{-2} \text{s}^{-1}$) for 10 min; samples were treated with (+) or without (–) lambda phosphatase (λPPase). Samples were fractionated in 10% NuPAGE gels and immunoblots probed with anti-CRY2 antibodies.

was found previously that expression of a GUS–CCT2 (CRY2 C-terminal domain) fusion protein caused a constitutive photomorphogenic phenotype in transgenic plants. For example, transgenic seedlings expressing GUS–CCT2 developed short hypocotyls in the dark or in the light, whereas wild-type plants exhibit inhibition of hypocotyl elongation only in response to light²⁸. This observation indicates that a light-induced biochemical change may occur in the carboxy-terminal domain of *cry2*, resulting in activation of the photoreceptor²⁸.

We hypothesized that *cry2* phosphorylation may occur in its C-terminal domain, in which several putative protein phosphorylation motifs are found (not shown). We further reasoned that the GUS–CCT2 fusion protein, which contains no chromophore-binding domain but is constitutively active, might show a constitutive phosphorylation regardless of light treatment. We tested this hypothesis by analysing phosphorylation of GUS–CCT2 fusion protein in etiolated or blue-light-treated transgenic plants expressing GUS–CCT2, using the *in vivo* ³²P-labelling immunoprecipitation assay (Fig. 5). In contrast to the endogenous *cry2* that was phosphorylated only in blue-light-treated seedlings (Fig. 1a), the GUS–CCT2 fusion protein, immunoprecipitated by anti-GUS antibody, was labelled with ³²P in both dark-grown and blue-light-treated plants (Fig. 5). Residues of the GUS–CCT2 fusion protein that were phosphorylated must be in the CCT2 moiety because no phosphorylation was detected in GUS protein immunoprecipitated from transgenic plants expressing GUS (Fig. 5). This result indicates that the constitutive light response found in transgenic plants expressing the GUS–CCT2 fusion protein was due to a constitutive phosphorylation in the *cry2* C-terminal domain of the fusion

protein and that *cry2* activity is dependent on the phosphorylation of its C-terminal domain.

Our results demonstrated that the function and turnover of *Arabidopsis* cryptochrome 2 is regulated by the blue-light-dependent phosphorylation. Because different cryptochromes have similar structures¹, it is likely that the blue-light-dependent phosphorylation is also important for the function of cryptochromes in other organisms. We propose that absorption of photons by a cryptochrome changes its conformation, enabling phosphorylation of the photoreceptor; the phosphorylated cryptochrome triggers signal transduction and physiological responses. The phosphorylation of a cryptochrome also marks it for degradation. The light-induced degradation of the active form of a cryptochrome may serve to regulate the cryptochrome activity in the presence of light and to de-sensitize the photoreceptor in the absence of light. □

Methods

Plant materials

Arabidopsis mutants and transgenic lines including *cry2-1* (ref. 6), GUS–CCT2 (ref. 28), *det1-1* (ref. 19), *cop1-4* and *cop1-6* (ref. 14), and the phytochrome mutants (*phyA-201*, *phyB-1*, *phyD-1*, *phyE-1*, *phyAB*) are as described²⁵ except for *phyABD* and *phyBDE* (G.C.W., unpublished work). The CRY2-overexpressing transgenic line in Col-4 background (CRY2+) was prepared as described⁵. Seeds treated with or without bleach-sterilization were sown on MS medium (Sigma, M-9274) or in soil, respectively, kept in the dark at 4 °C for 3 days, exposed to white light for 4 h to promote germination, and grown in the dark for 5–7 days before use. Light sources are as described²⁴.

Immunoprecipitation and immunoblot analyses

For immunoprecipitation, seedlings were excised from the hypocotyl base under a dim red safe light ($<5 \mu\text{mol m}^{-2} \text{s}^{-1}$), and homogenized in ice-cold immunoprecipitation buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 10% glycerol, 0.2% Triton X-100, 10 μM NaF, 1 mM PMSF), and 1 × complete protease inhibitor cocktail (Roche). Extracts were passed through 0.2-μm filters. Antibodies were added (1:200) to the filtrate, incubated in ice for 1 h, protein-A-sepharose (Sigma) was then added (1:60) to the mixture and incubated in ice for 1 h. Immunoprecipitation complex was washed three times with ice-cold washing buffer (50 mM Tris-HCl pH 8, 140 mM NaCl, 1 mM EDTA, 0.1% Triton X-100), mixed with SDS–PAGE sample buffer, and boiled for 5 min.

Proteins were separated either in conventional 10% Laemmli gels using Tris-glycine buffer or in precast 10% Bis-Tris NuPAGE gels (NuPAGE NP0303, Invitrogen) using MOPS buffer according to the instructions of the manufacturer. NuPAGE gels were run at a constant voltage (100 V) at 13 °C for 6–7 h. Proteins were blotted to nitrocellulose membranes for autoradiographs or immunoblots. Immunoblots were probed with primary antibodies, and then with goat-anti-rabbit IgG conjugated with horseradish peroxidase (Amersham), and developed using the enhanced chemiluminescence (ECL) method as described previously²⁹. Protein signals in an immunoblot were scanned, digitized and quantified using NIH Image. Different immunoblots are not directly comparable, because the exposure time of ECL cannot be precisely controlled between different experiments.

Phosphorylation assay

For *in vivo* labelling analyses, six-day-old, dark-grown seedlings (about 30) were cut from the hypocotyl base, incubated with ³²P (H_3PO_4 300 μCi, ICN) in water for 2 h in the dark at room temperature. The excess ³²P was removed by rinsing in water. Plantlets were treated under different light conditions before being collected for immunoprecipitation

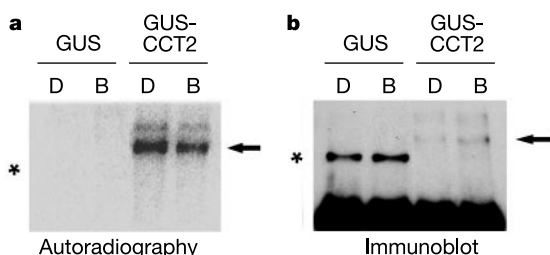


Figure 5 Constitutive phosphorylation of the CRY2 C-terminal domain moiety (CCT2) of the GUS–CCT2 fusion protein. The autoradiographs (**a**) and immunoblots (**b**) were prepared from the same nitrocellulose membrane, using the method as described in Fig. 1 except that the immunoprecipitation and immunoblot were reacted with anti-GUS antibodies. Transgenic plants expressing GUS (GUS) or GUS–CCT2 (GUS–CCT2)²⁸ were treated with (B) or without (D) blue light ($5 \mu\text{mol m}^{-2} \text{s}^{-1}$) for 20 min before harvesting. Arrows and asterisks indicate GUS–CCT2 and GUS, respectively.

and immunoblot analyses. For lambda phosphatase treatment, seedlings were homogenized in lambda phosphatase buffer provided by the manufacturer (New England Biolabs), and divided into two parts. One part was incubated without lambda phosphatase, the other part was incubated with 400 units of lambda phosphatase (New England Biolabs) at 30 °C for 30 min. The reactions were stopped by the addition of SDS-PAGE sample buffer, boiled and analysed by immunoblots as described.

Received 11 January; accepted 27 March 2002; doi:10.1038/nature00815.

1. Cashmore, A. R., Jarillo, J. A., Wu, Y. J. & Liu, D. Cryptochromes: blue light receptors for plants and animals. *Science* **284**, 760–765 (1999).
2. Sancar, A. Cryptochrome: the second photoactive pigment in the eye and its role in circadian photoreception. *Annu. Rev. Biochem.* **69**, 31–67 (2000).
3. Holmes, M. G. (ed.) *Photoreceptor Evolution and Function* (Academic, New York, 1991).
4. Briggs, W. R. & Huala, E. Blue-light photoreceptors in higher plants. *Annu. Rev. Cell Dev. Biol.* **15**, 33–62 (1999).
5. Lin, C. *et al.* Enhancement of blue-light sensitivity of *Arabidopsis* seedlings by a blue light receptor cryptochrome 2. *Proc. Natl Acad. Sci. USA* **95**, 2686–2690 (1998).
6. Guo, H., Yang, H., Mockler, T. C. & Lin, C. Regulation of flowering time by *Arabidopsis* photoreceptors. *Science* **279**, 1360–1363 (1998).
7. Ahmad, M., Jarillo, J. A. & Cashmore, A. R. Chimeric proteins between cry1 and cry2 *Arabidopsis* blue light photoreceptors indicate overlapping functions and varying protein stability. *Plant Cell* **10**, 197–208 (1998).
8. El-Din El-Assal, S., Alonso-Blanco, C., Peeters, A. J., Raz, V. & Koornneef, M. A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genet.* **29**, 435–440 (2001).
9. Mas, P., Devlin, P. F., Panda, S. & Kay, S. A. Functional interaction of phytochrome B and cryptochrome 2. *Nature* **408**, 207–211 (2000).
10. Wang, H., Ma, L. G., Li, J. M., Zhao, H. Y. & Deng, X. W. Direct interaction of *Arabidopsis* cryptochromes with COP1 in light control development. *Science* **294**, 154–158 (2001).
11. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
12. Guo, H., Duong, H., Ma, N. & Lin, C. The *Arabidopsis* blue light receptor cryptochrome 2 is a nuclear protein regulated by a blue light-dependent post-transcriptional mechanism. *Plant J.* **19**, 279–287 (1999).
13. Fankhauser, C. & Chory, J. Light control of plant development. *Annu. Rev. Cell. Dev. Biol.* **13**, 203–229 (1997).
14. Kwok, S. F., Piekos, B., Misera, S. & Deng, X. W. A complement of ten essential and pleiotropic *Arabidopsis* COP/DET/FUS genes is necessary for repression of photomorphogenesis in darkness. *Plant Physiol.* **110**, 731–742 (1996).
15. Buche, C., Poppe, C., Schäfer, E. & Kretsch, T. *eid1*: A new *Arabidopsis* mutant hypersensitive in phytochrome A-dependent high-irradiance responses. *Plant Cell* **12**, 547–558 (2000).
16. Guo, H., Mockler, T. C., Duong, H. & Lin, C. SUB1, an *Arabidopsis* Ca²⁺-binding protein involved in cryptochrome and phytochrome coaction. *Science* **291**, 487–490 (2001).
17. Deng, X.-W., Caspar, T. & Quail, P. H. *COP1*: a regulatory locus involved in light-controlled development and gene expression in *Arabidopsis*. *Genes Dev.* **5**, 1172–1182 (1989).
18. McNellis, T. W. *et al.* Genetic and molecular analysis of an allelic series of cop1 mutants suggests functional roles for the multiple protein domains. *Plant Cell* **6**, 487–500 (1994).
19. Chory, J., Peto, C., Feinbaum, R., Pratt, L. & Ausubel, F. *Arabidopsis thaliana* mutant that develops as a light-grown plant in the absence of light. *Cell* **58**, 991–999 (1989).
20. Osterlund, M. T., Hardtke, C. S., Wei, N. & Deng, X. W. Targeted destabilization of HY5 during light-regulated development of *Arabidopsis*. *Nature* **405**, 462–466 (2000).
21. Yang, H. Q., Tang, R. H. & Cashmore, A. R. The signalling mechanism of *Arabidopsis* CRY1 involves direct interaction with COP1. *Plant Cell* **13**, 2573–2587 (2001).
22. Yeh, K. C. & Lagarias, J. C. Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry. *Proc. Natl Acad. Sci. USA* **95**, 13976–13981 (1998).
23. Ahmad, M., Jarillo, J. A., Smirnova, O. & Cashmore, A. R. The CRY1 blue light photoreceptor of *Arabidopsis* interacts with phytochrome A *in vitro*. *Mol. Cell* **1**, 939–948 (1998).
24. Mockler, T. C., Guo, H., Yang, H., Duong, H. & Lin, C. Antagonistic actions of *Arabidopsis* cryptochromes and phytochrome B in the regulation of floral induction. *Development* **126**, 2073–2082 (1999).
25. Devlin, P. F. *et al.* Phytochrome D acts in the shade-avoidance syndrome in *Arabidopsis* by controlling elongation growth and flowering time. *Plant Physiol.* **119**, 909–915 (1999).
26. Devlin, P. F., Patel, S. R. & Whitelam, G. C. Phytochrome E influences internode elongation and flowering time in *Arabidopsis*. *Plant Cell* **10**, 1479–1488 (1999).
27. Parks, B. M. & Quail, P. H. Phytochrome-deficient *hy1* and *hy2* long hypocotyl mutants of *Arabidopsis* are deficient in chromophore biosynthesis. *Plant Cell* **3**, 1177–1186 (1991).
28. Yang, H.-Q. *et al.* The C termini of *Arabidopsis* cryptochromes mediate a constitutive light response. *Cell* **103**, 815–827 (2000).
29. Lin, C., Ahmad, M. & Cashmore, A. R. *Arabidopsis* cryptochrome 1 is a soluble protein mediating blue light-dependent regulation of plant growth and development. *Plant J.* **10**, 893–902 (1996).

Acknowledgements

We thank T. Cashmore, X.-W. Deng, J. Chory and T. Kretsch for providing *Arabidopsis* mutant lines. This work is supported by research grants from the National Institutes of Health, the National Science Foundation and UCLA-FGP. T.C.M. is supported by a predoctoral UCLA-NSF/IGERT training award.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.L. (e-mail: clin@mcdub.ucla.edu).

Structure of the SRP19–RNA complex and implications for signal recognition particle assembly

Tobias Hainzl, Shenghua Huang & A. Elisabeth Sauer-Eriksson

Umeå Centre for Molecular Pathogenesis, Umeå University, SE-901 87 Umeå, Sweden

The signal recognition particle (SRP) is a phylogenetically conserved ribonucleoprotein. It associates with ribosomes to mediate co-translational targeting of membrane and secretory proteins to biological membranes. In mammalian cells, the SRP consists of a 7S RNA and six protein components. The S domain of SRP comprises the 7S.S part of RNA bound to SRP19, SRP54 and the SRP68/72 heterodimer; SRP54 has the main role in recognizing signal sequences of nascent polypeptide chains and docking SRP to its receptor^{1–3}. During assembly of the SRP, binding of SRP19 precedes and promotes the association of SRP54 (refs 4, 5). Here we report the crystal structure at 2.3 Å resolution of the complex formed between 7S.S RNA and SRP19 in the archaeon *Methanococcus jannaschii*. SRP19 bridges the tips of helices 6 and 8 of 7S.S RNA by forming an extensive network of direct protein–RNA interactions. Helices 6 and 8 pack side by side; tertiary RNA interactions, which also involve the strictly conserved tetraloop bases, stabilize helix 8 in a conformation competent for SRP54 binding. The structure explains the role of SRP19 and provides a molecular framework for SRP54 binding and SRP assembly in Eukarya and Archaea.

The secondary structures of archaeal 7S RNAs are highly similar to their mammalian counterparts. These RNAs of about 300 nucleotides contain eight helical elements, which fold into the *Alu* and S domains connected by a long linker. Analyses of genome sequences indicate that archaeal species code for only two homologues of the six eukaryotic SRP proteins, namely SRP19 and SRP54 (ref. 6). In eubacteria, a minimal set of SRP subunits exists, and this is composed of 4.5S RNA, a single RNA helix with a region homologous to helix 8, and the SRP54 homologue Ffh.

SRP biogenesis in higher eukaryotes involves the sequential binding of SRP19 and SRP54 proteins to 7S RNA^{4,5}. Archaeal SRP54 has an intrinsic affinity for 7S RNA *in vitro*, yet the presence of SRP19 significantly enhances SRP54 attachment⁷. The SRP19 binding site is situated in the apices of helices 6 and 8 of 7S RNA^{8,9}. Binding of SRP19 leads to a restructuring of both helices, causing localized changes at the SRP54 binding site, which comprises the symmetric and asymmetric internal loops of helix 8 (refs 10, 11). Although crystal structures for helix 6 in complex with SRP19 (ref. 12) and helix 8 in complex with the M domain of Ffh¹³ have been described, the protein–RNA and RNA–RNA interactions within the complete S domain of SRP were still unknown. We therefore determined the crystal structure of SRP19 bound to the S domain of RNA from *M. jannaschii*. Co-crystals of wild-type SRP19 in complex with a 97-nucleotide 7S.S RNA fragment, corresponding to helices 6 and 8, the U-turn, and part of helix 5, diffracted to 2.3 Å resolution and led to structure determination by molecular replacement (Fig. 1).

In the 7S.S RNA–SRP19 complex, RNA helices 6 and 8 associate in a parallel, slightly tilted mode. Helix 6 forms an overall rigid stem-loop structure with continuous base stacking (except for two looped-out adenosines, A176 and A177) and base pairing of canonical- or wobble-type throughout. In contrast, base mismatches in two regions of helix 8 cause a severe distortion from regular A-form RNA. Within the symmetric loop, bases C201–G205 and A216–A220 form five consecutive noncanonical base pairs.

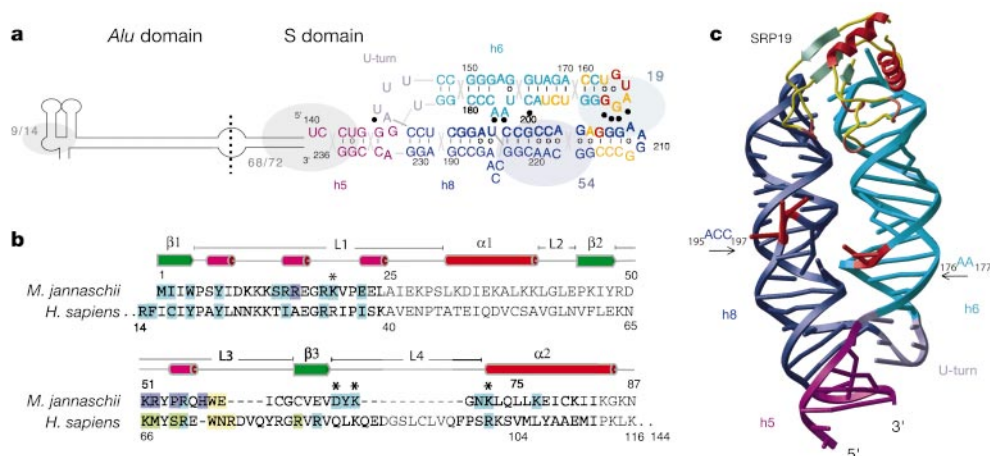


Figure 1 Structure of the 7S.S RNA in complex with SRP19. **a**, Secondary structure of 7S RNA in Archaea and Eukarya with the approximate positions of the binding sites for SRP proteins indicated. Archaea lack genes for both heterodimers, SRP9/14 and SRP68/72. The nucleotide sequence and two-dimensional topology diagram of the 7S.S RNA fragment used for crystallization are given, with the numbering corresponding to full-length *M. jannaschii* 7S RNA. Lines between bases indicate WC base pairs, open circles non-canonical pairs and filled circles tertiary RNA–RNA interactions. Direct contacts between the RNA backbone and SRP19 are orange, base-specific contacts in red. **b**, Sequence alignment of *M. jannaschii* and human SRP19 on the basis of structural

superposition²⁵. The secondary structural assignment is based on *M. jannaschii* SRP19 and the positions of two α -helices, four 3_{10} -turns and three β -strands are indicated. Amino acids in contact with helices 6 and 8 are boxed in cyan and blue. Amino acids in human SRP19 in contact with helix 6 are taken from 1JID (ref. 12). Four residues in *M. jannaschii* SRP19 that form side-chain-specific contacts are indicated with asterisks. Model-based contacts between SRP19 and SRP54 are shown in yellow, and between human SRP19 and helix 8 in green. **c**, Three-dimensional ribbon representation of *M. jannaschii* 7S.S RNA bound to SRP19 with the RNA helical elements coloured as in **a**.

Furthermore, A194•U224 and G193•A225 form mismatches proximal to three looped-out nucleotides, A195–C197. Helix 8 stacks coaxially onto helix 5, forming a continuous helical subdomain that connects to helix 6 via a U-turn, a motif also present in the Alu domain¹⁴.

Helices 6 and 8 are clamped together by SRP19 and the tertiary RNA interactions that occur at two major sites. One site comprises the tetraloop–tetraloop interface, where contacts are primarily made by the RNA backbones but also include important base–base contacts of the strictly conserved nucleotides in the tetraloops of both helices. Tetraloops are common and important structural elements in large RNAs and they are thought to serve as nucleation sites for RNA folding and to stabilize RNA tertiary structures¹⁵. The helix 8 tetraloop (209–GGAA–212) belongs to the most abundant GNRA-type (where N is any nucleotide, and R is A or G), whereas helix 6 is capped by an unusual GNAR-type tetraloop (163–GUAG–166). The sequences of both loop-types are conserved in all 7S RNAs, and represent the only conserved region in helix 6 (ref. 6). The GGAA sequence has the previously described compact fold^{13,16}, whereas the structure of the GUAG loop deviates from known GNAR-type tetraloops, which also includes the GGAG loop of the human homologue¹². In the human structure, the last three bases of the tetraloop are stacked, whereas in *M. jannaschii* the second base, U164, is looped out. Moreover, the archaeal tetraloop adopts a more closed conformation, and a direct nucleotide interaction is formed between the first base, G163, and the fourth base, G166, that is water-mediated in the human structure. The stacked third and fourth bases of the helix 6 tetraloop in *M. jannaschii* interact with bases in the minor groove of helix 8. The invariant A165 participates in a Watson–Crick (WC)-type base triple with the strictly conserved fourth adenine of the tetraloop A212•G209 pair, whereas G166 is part of a quadruple-base interaction with G163 and the guanine of the tetraloop-closing G213•C208 pair (Fig. 2a). These long-range RNA interactions provide a molecular rationale for prior mutagenesis results, which showed that base substitutions at positions 3 and 4 of the helix 6 tetraloop destabilize the SRP19–7S.S RNA complex⁸, and highlight the importance of these triples for the integrity of the 7S.S RNA tertiary structure.

The second set of RNA–RNA interactions links helix 6 with the asymmetric loop of helix 8. Results from chemical modification identified this loop as a critical region for SRP19-induced conformational changes^{10,11}. In the *Escherichia coli* helix 8–Ffh complex, three bases of the asymmetric loop produce a platform onto which SRP54 docks¹³. In the homologue region of *M. jannaschii*, nucleotides A195–C197 protrude from the helix 8 duplex, whereas the opposing strand is continuously base paired, thus defining this loop as a bulge loop (Fig. 2b). The phosphoribose backbone in the bulge loop is closely packed, and with the backbone inverted, the bases are exposed on the surface and positioned in a plane almost perpendicular to the helical axis. Elevated temperature factors of the nucleotides in the bulge loop indicate some degree of mobility. Three riboses of the base pairs flanking the bulge loop (C198•G223 and A194•U224) are locked into place by direct RNA interactions with helix 6. U224 and C198 bind to the looped-out adenosines in helix 6, A176 and A177, respectively, whereas G223 is stabilized by a hydrogen bond to C178 (Fig. 2b).

In agreement with the biochemical evidence^{10,11}, helix 8 in the

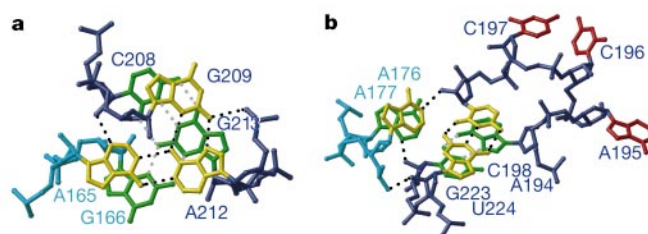


Figure 2 Geometries of RNA–RNA interactions at the tetraloop and the bulge loop regions. **a**, Long-range base contacts between the tetraloops showing the two minor-groove base triples. A165, A212 and G209 are highlighted in yellow and G166, G213 and C208 in green. The phosphoribose backbones of helices 6 and 8 are depicted in cyan and blue, respectively. **b**, Tertiary RNA contacts at the bulge loop of helix 8. A176 stacks beneath A177 and both bases loop into the minor groove of helix 8 and form bonds with 2'-OH of the loop-flanking nucleotides U224 and C198. The surface-exposed bases of the nucleotides in the bulge loop, A195–C197, are shown in red.

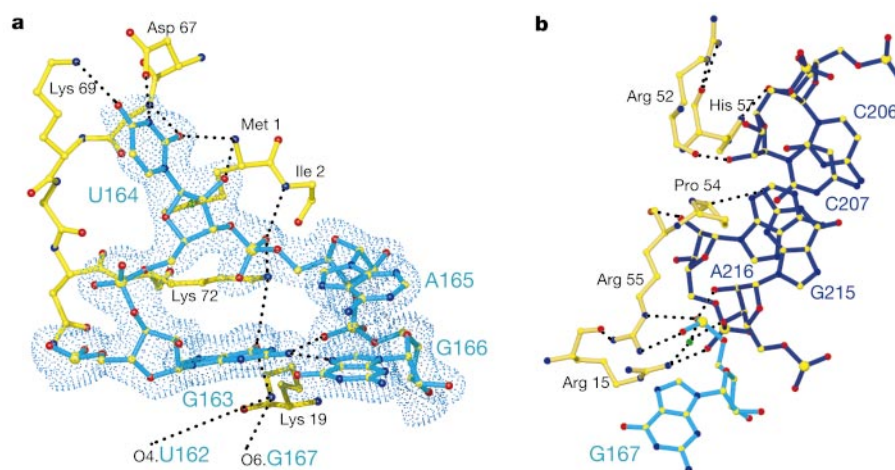


Figure 3 Details of the RNA–SRP19 interactions in *M. jannaschii*. **a**, The interactions between SRP19 and helix 6 in a view emphasizing two major contact areas. The looped-out base U164 makes pseudo-WC interactions with the side chains of Asp 67, Tyr 68 and Lys 69, and a large hydrogen-bond network includes contacts from residues Ile 2, Lys 72 and Lys 19 to the backbone of U162, the base of G163, and to both bases of the closing

U162•G167 pair. The difference electron density map ($2|F_o| - |F_c|$) is shown over the helix 6 tetraloop and contoured at 1σ . **b**, SRP19 interactions with helix 8. The conserved residue Arg 55 has a key position bonding to backbone oxygen atoms of both RNA helices (G167 and A216). The water molecule W304 (green) also bridges both RNA helices to form hydrogen bonds to the backbone oxygens of G167 and G215.

SRP19–7S.S complex of *M. jannaschii* corresponds to the conformation recognized by *E. coli* Ffh. The distal four conserved mismatches in the symmetric loop have a base-pairing scheme identical to that seen in the crystal structures of the *E. coli* homologue^{13,17}. Significantly, the relative orientation of the symmetric loop to the proximal stem and bulge loop, as well as the perpendicular array of the nucleosides in the bulge loop, are strikingly similar to the eubacterial helix 8 when bound to SRP54 (ref. 13). The exposed bases in the bulge loop are most probably stabilized by the stacked configuration, which forms upon SRP54 binding¹³. Notably, the WC edges of adenosines A194 and A225, which constitute a cross-strand A-stack¹⁸ proximal to the bulge loop, are displayed as a flattened minor groove free for potential recognition by the NG domain of SRP54.

The *M. jannaschii* SRP19, like its human counterpart¹², is an α/β -type protein comprising a core of three antiparallel β -strands packed against two α -helices and four 3_{10} -turns situated at two extended loop regions, L1 and L3 (Fig. 1). The proteins superimpose well, with a r.m.s. deviation of 1.1 Å (see Methods), and interact with a widened major groove at the tip of helix 6. Contacts are primarily formed with the phosphoribose backbone, but unlike in the human complex, four direct side-chain interactions with bases from the GUAG tetraloop and its closing U162•G167 wobble-pair are present in the *M. jannaschii* complex. The surface-exposed position of the looped-out U164 is stabilized by interactions with Tyr 68, Asp 67 and Lys 69, with the latter two residues forming base-specific contacts. In addition, the side chains of the highly conserved Lys 19 and Lys 72 form a large hydrogen-bonding network together with U162, G163 and G167 (Fig. 3a). The SRP19-binding site on helix 8 is situated in the minor groove of the distal stem. Contacts are made from residues in loop L3, where Lys 51, Arg 52 and His 57 interact with the backbone of C206–G209 in the 3'-strand, and Arg 55 with 2'-OH of A216 in the 5' strand. Moreover, Arg 15 of loop L1 binds to the phosphate oxygen of A216 (Fig. 3b). Two conserved residues, Pro 54 and Lys 51, occupy important positions. The carbonyl of Pro 54 forms a base-specific contact with G215, and its side chain is wedged between helices 6 and 8, whereas Lys 51 hydrogen bonds to G209 situated in the proposed signal sequence binding platform formed by the helix 8 backbone¹³. The SRP19–helix 8 interface is considerably smaller and covers about 800 Å² of solvent-accessible surface area, compared with about 1,700 Å² for

the SRP19–helix 6 interface, which is consistent with a primary binding site of SRP19 on helix 6 (ref. 8).

The human SRP19-binding site on helix 6 is formed by two separate regions: nucleotides G147–G151 in the tetraloop and C140–C143 in the distal stem¹². Unexpectedly, only nucleotides bound directly to human SRP19 superimpose well with nucleotides from *M. jannaschii* when the protein components are superimposed (Fig. 4a, b). However, whereas the GNAR tetraloop nucleotides 3 and 4 are identically positioned, the nucleotides in the distal stem are out of register by one base. Thus nucleotides C158 and U159 in *M. jannaschii* helix 6 structurally coincide with human C141 and U142, and not with U142 and C143 as predicted from sequence alignment (Fig. 4b). This base shift is due to an increased helical rise in the distal turn of the archaeal helix 6. The substantially shallower and more accessible major groove in *M. jannaschii* provides a more favourable SRP19-binding site involving a stretch of 12 nucleotides (U157–G168).

Comparing the human and archaeal complexes demonstrates that different helix 6 tertiary structures are able to display identical SRP19-recognition sites. This structural diversity may be intrinsic where tighter interactions are necessary in the archaeal complex for stability at extreme temperatures. But another interpretation based on the absence of helix 8 in the human complex is plausible. In time-resolved footprinting studies, it has been shown that the assembly of the SRP19–7S RNA complex involves structurally distinct intermediates¹¹. The helix 6 structures therefore may represent different RNA conformations that occur during SRP assembly. In support of this interpretation, the human complex contains a high proportion of water-mediated interactions with no base-specific contacts. It is possible that docking of helix 8 to the helix 6–SRP19 complex induces the conformational change in the RNA, which results in the observed base shift and tighter binding. Intriguingly, the side chain of the highly conserved Arg 18 residue (Arg 33 in the human) is optimally positioned in the claw-shaped loop L1 to mediate such a structural shift (Fig. 4d).

In *E. coli*, the distal and proximal helical regions of helix 8 adopt different relative orientations to each other depending on SRP54 binding, with the connecting asymmetric loop functioning as a hinge^{13,17}. The relative orientation of these regions in the *M. jannaschii* 7S.S RNA–SRP19 complex is similar to the ones found in *E. coli* helix 8 bound to SRP54, whereas an overlay of the proximal

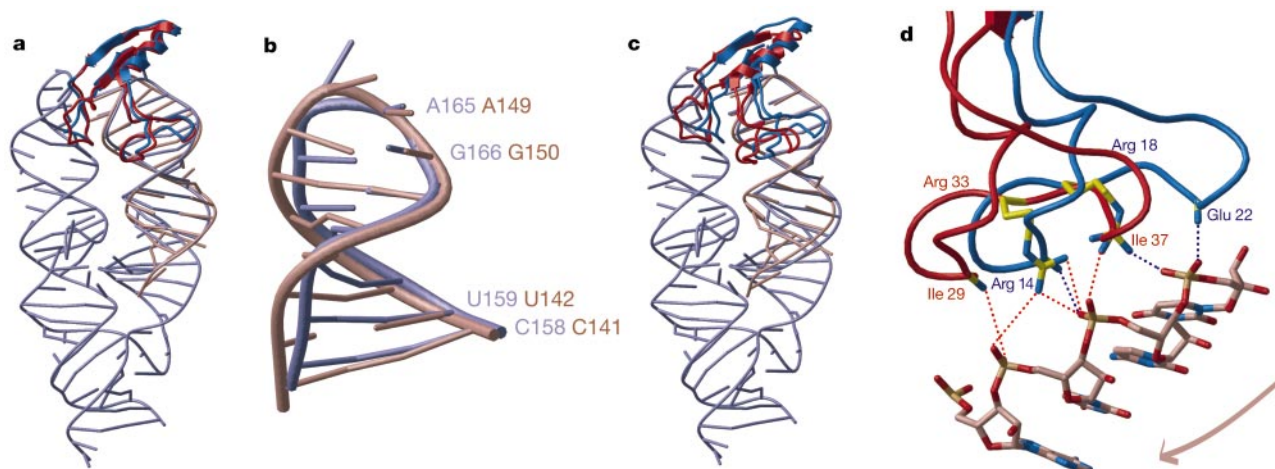


Figure 4 Structural differences between helix 6 from human and *M. jannaschii* SRP RNA. **a**, Structural alignment of the protein components showing the good structural fit of the two proteins. Only residues 16–84 in humans and 1–66 in *M. jannaschii* are shown for clarity. RNA and SRP19 from *M. jannaschii* are shown in light and dark blue, and their human counterparts are shown in light and dark red (coordinates are from 1JID¹²). **b**, Close-up view of the RNA helices in **a**. The helices do not match except for the nucleotides defining the two protein-binding regions of human helix 6. **c**, Structural alignment of helix 6 based on 16 sequential phosphate atoms at the apex of helix 6 (U157–A172 in *M. jannaschii* and C141–G156 in humans) showing the large deviation of

the SRP19 proteins with C α –C α shifts of up to 6 Å for residues in the L1 RNA-binding loop. **d**, Close-up view of the protein–RNA interactions in **c**. Despite the large structural deviations, identical protein interactions are made to the phosphoribose backbone of C158 and U159 in *M. jannaschii* and C141 and U142 in humans. RNA binding in the claw-like fold of loop L1 involves the main-chain nitrogen atoms of Arg 14 and Glu 22 (Ile 29 and Ile 37 in humans) and side-chain nitrogens of the conserved residue Arg 18 (Arg 33). The guanido group of Arg 18 (Arg 33) bridges oxygens from neighbouring phosphate groups and seems ideally positioned to aid an eventual RNA slide.

helices of *M. jannaschii* and the free form of *E. coli* positions the distal region of the *E. coli* helix 8 apart from helix 6. Assuming an analogous relative orientation of the proximal and distal regions in helix 8 of free 7S RNA, these findings suggest that the main role of SRP19 is to properly position these regions. The concerted interactions between helix 6–SRP19 and helix 8, pulling the tetraloops together while constraining the mobility of the proximal helix, enforce a reorientation of the symmetric loop and induce a bulge-

loop geometry. Cumulatively, this leads to high-affinity SRP54 binding. The proposed mechanism is similar to the one described for S15 binding to 16S rRNA and may reflect a general function of primary RNA-binding proteins in the assembly of large RNPs¹⁹.

The SRP54M domain is modelled onto the 7S.S RNA–SRP19 complex without causing steric clashes. In the model of the ternary S domain RNP, the minor groove of helix 8 forms a contiguous recognition surface for SRP19 and SRP54 (Fig. 5a). The proximity of their main chains in the protein–protein interface (6.5–8 Å) suggests that contacts from side chain to side chain may occur between residues Trp 72–Arg 74 in loop L3 of SRP19 and residues Arg 408–Ser 413 within the helix–turn–helix (HTH) motif of SRP54 (Fig. 5b). Prior biochemical studies were unable to detect direct interaction of the free proteins, yet structuring of the SRP19–SRP54 interface upon 7S RNA binding and subsequent interaction between the bound proteins might be an important factor in completing the SRP assembly. Further structural studies of the ternary SRP19–SRP54–7S RNA complex are needed for the complete understanding of the principles of SRP assembly in archaea and human. □

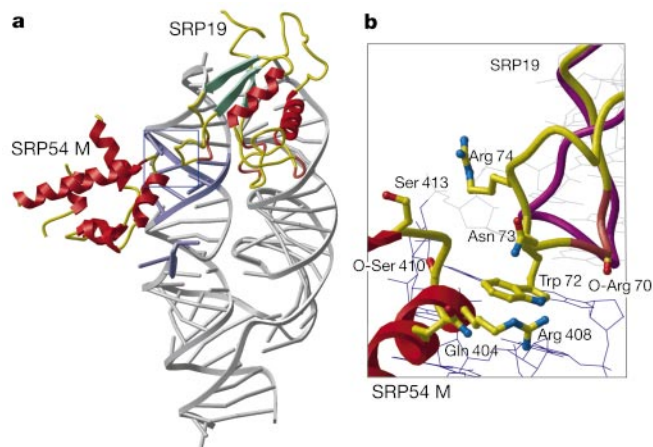


Figure 5 Model of human SRP19 and SRP54 M-domain proteins bound to *M. jannaschii* 7S.S RNA. **a**, The binding site of the human SRP54M domain was obtained by superimposing C199–G222 with the homologous region in the helix 8–Ffh M-domain complex (1DUL¹³) and substituting the *E. coli* Ffh M domain and *M. jannaschii* SRP19 with their human counterparts (1QB2²⁹ and 1JID¹²). Nucleotides of the symmetric and bulge loops in helix 8 are coloured in blue. **b**, Enlarged view of the modelled interface between loop L3 of SRP19 and the HTH motif of SRP54. Backbone atoms of the noncanonical G•A pair in the symmetric loop make contact to both proteins: that is, with SRP19 in the 3′-strand (A216 with Arg 70) and with SRP54 in the 5′-strand (G205 with Ser 410). The C α tracing of SRP19 from *M. jannaschii* is shown in magenta. The model is based on a rigid docking of the protein molecules onto helix 8.

Methods

Preparation of SRP19–RNA complex

The 97-nucleotide RNA fragment corresponding to nucleotides U140–C236 of *M. jannaschii* 7S RNA was transcribed *in vitro* and purified as described²⁰. The *srp19* gene was amplified by PCR from methanococcal genomic DNA and the wild-type sequence cloned into the pET-3a vector (Novagen), and expressed in BL21-CodonPlus(DE3) cells (Stratagene). The SRP19 protein was purified by Mono S, Heparin Sepharose and Superdex 75 (Pharmacia) and stored in aliquots at –80 °C. Before SRP19–RNA complex formation, the 7S RNA was annealed in water by denaturation at 80 °C followed by snap cooling on ice. The complex was reconstituted at room temperature in a buffer containing 20 mM Tris–HCl at pH 8.0, 250 mM KCl, 5 mM MgCl₂, 1% β -mercaptoethanol and 5% glycerol, and purified on Mono Q (Pharmacia).

Complex formation, crystallization and data collection

The SRP19–RNA complex (3 mg ml^{–1}) was crystallized by the hanging-drop vapour-diffusion technique at 18 °C. Crystals (0.15 × 0.1 × 0.05 mm³) grew in 5 d when the protein solution was mixed with an equal volume of mother liquor containing 18% polyethylene glycol (PEG) 3350, 0.1 M HEPES buffer at pH 7.0, 0.1 M MgCH₃COO, 0.02 M CaCl₂, 5% dimethylsulphoxide and equilibrated against the same solution. For data collection, a single crystal was supported in a cryo loop instantly cooled at 100 K in a nitrogen stream, and diffraction data were measured using synchrotron radiation

($\lambda = 1.12 \text{ \AA}$) at beam line I711 MAX lab. Crystals belong to space group $P2_12_12_1$ ($a = 42.5 \text{ \AA}$, $b = 61.4 \text{ \AA}$, $c = 156.0 \text{ \AA}$) and contain one molecule in the asymmetric unit. Diffraction data from one crystal (98.1% complete to $2.3 \text{ \AA}$ resolution) were processed using the program DENZO²¹. The 78,892 independent observations were scaled and merged with SCALA²², yielding 18,606 unique reflections in the $20\text{--}2.3 \text{ \AA}$ resolution range, with an R_{merge} of 9.9% (the $2.38\text{--}2.30 \text{ \AA}$ shell is 91% complete with an R_{merge} of 33.2%).

Phasing and refinement

Truncated models of helix 6 (nucleotides 154–162 and 167–175) and helix 8 (nucleotides 199–222) derived from the human structure (Protein Data Bank accession code 1D4R²³) and the *E. coli* structure (1DUL¹³), respectively, were used in molecular replacement searches using program CNS²⁴. A combination of searches including the RNA helices either individually or together using X-ray data from 9.0 to $3.0 \text{ \AA}$ resolution yielded the correct solution with an initial R -factor of 43%. Additional nucleotides could be modelled into the resulting map and the position of the protein identified. Model building using program O²⁵ was followed by refinement using both CNS and REFMAC²². Five per cent of the data (955 reflections) randomly selected were excluded and used to calculate R_{free} to monitor the progress of refinement²⁶. The procedure of refinement and model building was iterated several times until the entire RNA molecule and SRP19 were traced. The refined structure has an R -factor of 22.5% (R_{free} 27.5%) using 17,651 independent reflections between 20 and $2.3 \text{ \AA}$ resolution. The electron density for nucleotides 183–188, 195–197 and 230–233 is not as strong as for the rest of the complex, but clearly shows their positions with the exception of nucleotide 184 in the U-turn, for which only the position of the base is defined. The r.m.s. deviations of the final model from ideal stereochemistry are $0.011 \text{ \AA}$ for bond lengths and 2.70° for bond angles. Five water molecules (W300–W304) that refine to low B -factors indicate the presence of metal ions. The r.m.s. deviation between human and *M. jannaschii* SRP19 proteins is based on C α atoms of residues Met 1–Gln 56, Cys 61–Tyr 68 and Asn 71–Ile 83 of *M. jannaschii* SRP19. Solvent-accessible surfaces were calculated with AREAIMOL²². Figures 1c, 2, 4 and 5 were prepared with ICM²⁷, and Fig. 3 with programs O²⁵ and Molray²⁸.

Received 5 March; accepted 17 May 2002; doi:10.1038/nature00768.

Published online 5 June 2002.

1. Lütcke, H. Signal recognition particle (SRP), a ubiquitous initiator of protein translocation. *Eur. J. Biochem.* **228**, 531–550 (1995).
2. Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. The signal recognition particle. *Annu. Rev. Biochem.* **70**, 755–775 (2001).
3. Wild, K., Weichenrieder, O., Strub, K., Sinning, I. & Cusack, S. Towards the structure of the mammalian signal recognition particle. *Curr. Opin. Struct. Biol.* **12**, 72–81 (2002).
4. Walter, P. & Blobel, G. Disassembly and reconstitution of signal recognition particle. *Cell* **34**, 525–533 (1983).
5. Politz, J. C. *et al.* Signal recognition particle components in the nucleolus. *Proc. Natl Acad. Sci. USA* **97**, 55–60 (2000).
6. Gorodkin, J., Knudsen, B., Zwieb, C. & Samuelsson, T. SRPDB (Signal Recognition Particle Database). *Nucleic Acids Res.* **29**, 169–170 (2001).
7. Bhuiyan, S. H., Gowda, K., Hotokezaka, H. & Zwieb, C. Assembly of archaeal signal recognition particle from recombinant components. *Nucleic Acids Res.* **28**, 1365–1373 (2000).
8. Zwieb, C. Recognition of a tetranucleotide loop of signal recognition particle RNA by protein SRP19. *J. Biol. Chem.* **267**, 15650–15656 (1992).
9. Zwieb, C. Site-directed mutagenesis of signal-recognition particle RNA. Identification of the nucleotides in helix 8 required for interaction with protein SRP19. *Eur. J. Biochem.* **222**, 885–890 (1994).
10. Diener, J. L. & Wilson, C. Role of SRP19 in assembly of the *Archaeoglobus fulgidus* signal recognition particle. *Biochemistry* **39**, 12862–12874 (2000).
11. Rose, M. A. & Weeks, K. M. Visualizing induced fit in early assembly of the human signal recognition particle. *Nature Struct. Biol.* **8**, 515–520 (2001).

12. Wild, K., Sinning, I. & Cusack, S. Crystal structure of an early protein–RNA assembly complex of the signal recognition particle. *Science* **294**, 598–601 (2001).
13. Batey, R. T., Rambo, R. P., Lucast, L., Rha, B. & Doudna, J. A. Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* **287**, 1232–1239 (2000).
14. Weichenrieder, O., Wild, K., Strub, K. & Cusack, S. Structure and assembly of the *Alu* domain of the mammalian signal recognition particle. *Nature* **408**, 167–173 (2000).
15. Uhlenbeck, O. C. Tetraloops and RNA folding. *Nature* **346**, 613–614 (1990).
16. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. & Pardi, A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **264**, 968–980 (1996).
17. Jovine, L. *et al.* Crystal structure of the fhl and EF-G binding sites in the conserved domain IV of *Escherichia coli* 4.5S RNA. *Struct. Fold. Des.* **8**, 527–540 (2000).
18. Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705–712 (1997).
19. Agalarov, S. C., Prasad, G. S., Funke, P. M., Stout, C. D. & Williamson, J. R. Structure of the S15, S6, S18-rRNA complex: Assembly of the 30S ribosome central domain. *Science* **288**, 107–112 (2000).
20. Price, S. R., Ito, N., Oubridge, C., Avis, J. M. & Nagai, K. Crystallization of RNA–protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249**, 398–408 (1995).
21. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
22. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
23. Wild, K., Weichenrieder, O., Leonard, G. A. & Cusack, S. The $2 \text{ \AA}$ structure of helix 6 of the human signal recognition particle RNA. *Struct. Fold. Des.* **7**, 1345–1352 (1999).
24. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
25. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
26. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–474 (1992).
27. Abagyan, R. A., Totrov, M. M. & Kuznetsov, D. N. ICM—a new method for protein modelling and design. Application to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488–506 (1994).
28. Harris, M. & Jones, T. A. Molray—a web interface between O and the POV-Ray ray tracer. *Acta Crystallogr. D* **57**, 1201–1203 (2001).
29. Clemons, W. M. Jr, Gowda, K., Black, S. D., Zwieb, C. & Ramakrishnan, V. Crystal structure of the conserved subdomain of human protein SRP54M at $2.1 \text{ \AA}$ resolution: evidence for the mechanism of signal peptide binding. *J. Mol. Biol.* **292**, 697–705 (1999).

Acknowledgements

We thank K. O. Stetter for providing *M. jannaschii* cells; C. Oubridge for T7 polymerase; and U. H. Sauer and T. Bergfors for suggestions to the manuscript. This work was supported by the Swedish Research Council and European Union. We thank K. Nagai and the Medical Research Council UK for their support at an early stage of this project.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to T.H. (e-mail: tobias.hainzl@ucmp.umu.se). Coordinates have been deposited in the Protein Data Bank under accession code 1LNG. Coordinates for the ternary SRP model have been deposited in the SRP database (SRPDB) and can be requested from T.H.

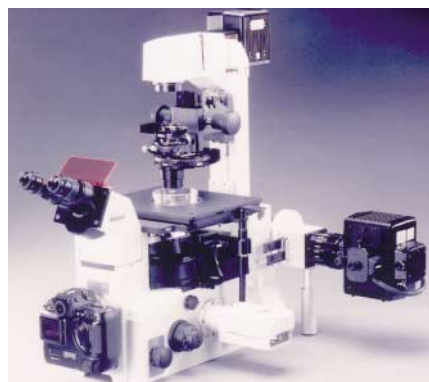
Microscopic detail

Microscopes and their accessories.

Eclipse TE2000

Nikon www.nikonusa.com
Inverted microscope for live-cell imaging

The TE2000 features a multi-port design and stratum structure for flexible configuration. Multiple optical accessories can be used, supporting multiple input illumination sources and output detector functions. Several epifluorescence techniques can be used simultaneously or sequentially, allowing for techniques such as high-resolution deconvolution 3D images, which may be incorporated together with laser illumination techniques. Three models are available: the basic TE2000S has two output ports, TE2000U with four output ports, an optional fifth user-defined position and an integrated magnification changer, and TE2000E, a high-precision Z-focus automated model with five output ports.



Live cells from all angles — Nikon's TE2000 series.

Spacers on the back of the metal frame ensure optimum distance to the knife-holder to facilitate sectioning without manual adjustments. All new Leica cryostats are equipped with the system for CE disposable blade knife-holders and CN steel knife-holders. Customers with existing CE and CN knife-holders can add the anti-roll system to their cryostats.



Microscopist's best friend: diamonds, that is.

It offers serial thick sections without chatter or compression and with no need to change knives in the middle of sectioning. The sections obtained are thinner and the diamond edge is durable and consistent. The knife is available in 4, 5, 6, 8 and 10 mm diamond widths. It can be used with hard and soft materials and works with all ultramicrotomes as well as regular microtomes with retractable knife-holders in the return phase.

i6000

BioGenex www.biogenex.com
All together now

Described by BioGenex as an 'all-in-one/all-at-once' instrument, the i6000 automated staining system can handle 60 slides and 60 reagents at up to 250 slides per 8-hour period. It can process special stains, immunohistochemistry, *in-situ* hybridization and haematoxylin-eosin slides in the same run. Results are standardized and the instrument operates in the bar-code mode for optimum workflow.

Cryostat anti-roll system

Leica www.leica-microsystems.com
Keep it in the frame

Leica's anti-roll system for cryomicrotomy consists of a reusable metal frame and a glass insert. All four edges of the insert can be used.

R-Gent SE-EM

Electron Microscopy Sciences

www.emsdiasum.com

A silver lining for electron microscopy

R-Gent SE-EM silver enhancement reagents are intended to enhance gold particles, particularly Aurion Ultra Small Particles, in EM visualization of immunodetection and *in-situ* hybridization experiments. The reagents are light-insensitive with low viscosity, high reproducibility, low autonucleation and time-controlled growth of particles.

BX61 microscope

Olympus America www.olympusamerica.com

Motorized microscope and software suite

The BX61 has a built-in, motorized Z-drive (0.01 μm resolution) and can be equipped with a motorized nosepiece, condenser, reflected light or fluorescence illuminator and up to three six-position filter wheels. Its MicroSuite software includes the ScopeView control module for quick changes between transmitted brightfield or darkfield, phase contrast, DIC, polarized, fluorescence and reflected brightfield or darkfield modes.

HISTO knife

Diatome www.emsdiasum.com

Through thick and thin

The HISTO knife is designed for thick sectioning of specimens between 0.2 and 10 μm .

Scanning stages

Prior Scientific

www.prior.com

Expanded range for advanced imaging

Prior Scientific has expanded its range of high-performance motorized scanning stages. Models are now available fitted with 1-mm pitch precision ground drive screws for finer resolution and superior repeatability. The stages are designed for more demanding imaging applications such as image tiling where very accurate positioning is required. A choice of 1 mm, 2 mm or 5 mm pitch drive screws is offered. The 2 mm stages are suitable for most routine positioning applications, while a stage fitted with 5 mm pitch drive screws is aimed at high-speed applications.

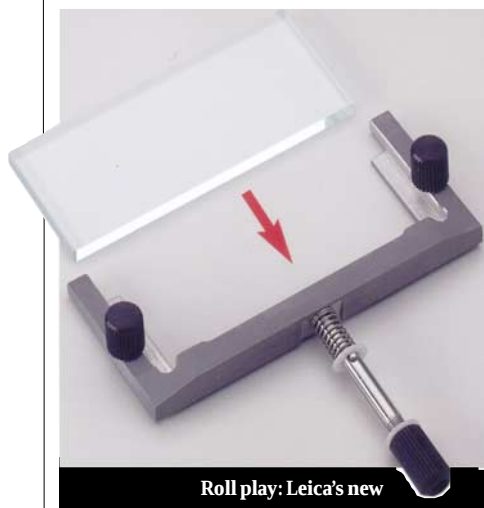
LyoVision

Biopharma Technology www.biopharma.co.uk

For film buffs

LyoVision is a new integrated suite of software for use with the Lyostat freeze-drying microscope system. It allows users to combine and display the images captured on camera with the temperature values, in the same file. This allows snapshots and movies of the freeze-drying process to be recorded and archived. The images can be saved as JPEGs with the user name and sample identity, date, time and temperature of the sample.

These notes are compiled in the Nature office from information provided by the manufacturers.



Contacts

Publisher: Fabien Savenay

Editor: Paul Smaglik

Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Senior European Sales Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Director: Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.johnson@naturejp.com

Asia-Pacific Advertising Manager:

Keyvin Johnson

naturejobs

A question of balance

The working group on women set up by the International Union of Pure and Applied Physics (IUPAP) met for the first time in Paris this spring and came up with 29 recommendations to promote sexual equality and to increase the number of women scientists (www.iupap.org). These recommendations were collected into seven categories, such as balancing career and family, getting women into leadership positions, and attracting more young females into the field.

But really, the list could have been split into two — discouraging discrimination and promoting affirmative action. Discouraging discrimination, by having a blind initial review of job applicants, for instance, is not too controversial. But promoting affirmative action — through quotas on women in tenure-track positions and on committees, for example — leaves many men and women uneasy.

But a report on sexual discrimination at the Massachusetts Institute of Technology (MIT), which came out at about the time that the IUPAP panel convened, indicates that such steps might be necessary — and not just in physics. The MIT report showed disparities in salary, high-level appointments and grant money from the school. In the wake of the report, MIT has made commitments to recruit more women and to appoint existing female faculty members to high-level positions. For example, the engineering department wants women in 20% of its faculty positions by 2012 (the figure reached 10% last year, up from 5% in 1990).

Other institutions seeking to follow MIT's lead, perhaps with an eye on IUPAP's list, would do well to pay close attention to some of its advice — try to present reforms as positive for both men and women. This is because the problem with affirmative action is not the good intentions, it is the resentment that pursuing such a goal can unintentionally generate.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

Teaching science graduates
to think business

p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS





Around the world, universities are starting to encourage entrepreneurship among science students. Steve Bunk investigates.

Tony Weiss, director of the Molecular Biotechnology Program that began last year at the University of Sydney, tells of a recent "information evening" to which prospective students were lured by drinks and nibbles to hear three talks. First, the managing director of an international biotech company detailed his diverse training. Then a technology-park executive waxed enthusiastic about challenges at the interface of business and industry. Finally, an academic told how "you can live a lot longer by working in a university".

Many of the students seemed willing to reject longevity in favour of a temporary university experience (including the drinks), followed by a ride aboard the biotech roller-coaster. "It's the buzz and excitement that draws in students," Weiss says.

His programme is at the forefront of an emerging trend: courses in science-based entrepreneurship. Similar courses — at the Karolinska Institute in Stockholm, Stanford University in California, and a cross-Atlantic partnership between the University of Cambridge, UK, and the Massachusetts Institute of Technology called the Cambridge-MIT Institute (CMI) — all rely on a mix of business and academic expertise. Funded by government and industry, they aspire to extend far beyond their own communities, sharing know-how at national and international levels.

Take the University of Sydney programme, which boasts external appointees at the Karolinska Institute and the National University of Singapore and is

planning student exchange programmes. Set up with more than A\$2 million (US\$1.1 million) from business and the federal government, the programme has enrolled 102 students in one undergraduate and three postgraduate molecular biotechnology degree courses. Industry partners are on the committee. "For a low investment they obtain gentle access to a high-powered academic in their field of interest," says Weiss.

TRANSATLANTIC PARTNERS

Business partners are also heavily involved in the CMI, but their activities are primarily outside education. The CMI, started in 2000 with a UK government grant of £65 million (US\$95 million) over five years, is the hub of a UK network linking industry with 50 university-based science enterprise centres established under a government initiative to commercialize science and technology. CMI private-sector partners participate in a variety of executive education and research programmes, and Cambridge and MIT staff conduct joint research and education projects.

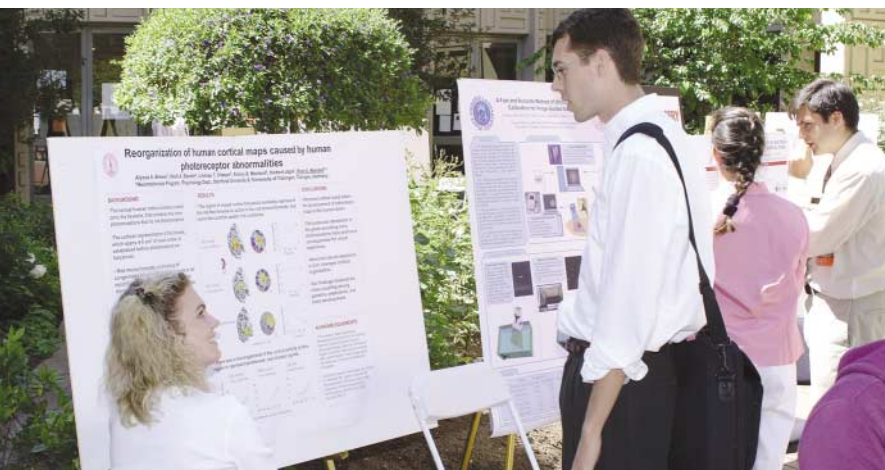
Three master's programmes are to be launched in October, in bioscience enterprise, technology policy, and environmental engineering and sustainable development. A course on chemical engineering will begin in the autumn of 2003.

Transatlantic, interactive lectures will be conducted using video links. For example, the bioscience enterprise course will offer information from Cambridge about the regulation of Britain's drug industry, while MIT provides details of the US system. "My dream is that we can have an international academic community," says Nick Oliver, director of the CMI's professional practice programme, which developed the courses.

The different teaching styles of the two universities provide "some interesting design challenges", Oliver says. MIT students complete weekly problem sets, with exams at the middle and end of each semester, whereas Cambridge has weekly supervisions with one exam at the end of the year. Both institutions have initiated independent evaluations of their methods to help them devise "hybrid" programmes, says David Good, director of the CMI's undergraduate education programme. Much anecdotal evidence is already coming from the student exchange programme, on which 60 students swapped places this year.

Another key aspect of the CMI effort is joint

Stanford University is encouraging science students to adopt a business-minded approach.



research projects involving the two universities — 28 so far. “Of course we anticipate a number of spin-offs and start-ups,” says Alan Windle, CMI executive director at Cambridge. As likely examples, he cites development of artificial bone material and ferro-electric memory devices for photonics systems. “But the CMI is a brand new form of partnership. We need new ways to gauge its overall impact,” he warns.

Other ways include generating appropriate models for technology transfer, laying the scientific base to drive new areas of technology (such as quantum computing) — and “a large investment” in educating business leaders on both sides of the Atlantic.

“One reason that the corporate world finds CMI attractive is the role CMI can play in putting them in contact with bright, highly educated students who have been exposed to both our cultures,” says John Vander Sande, the institute’s executive director at MIT.

GLOBAL LINKS

MIT also has alliances with two Singapore universities, as part of its goal to internationalize its education and research. The beginnings of a global network are visible.

Carl Johan Sundberg, a physician and physiology professor at Karolinska Institute, holds an external faculty appointment at the University of Sydney and leads a course in Sweden on science-based business enterprises. His course is a joint venture by Karolinska and three other schools that specialize in technology, economics, and arts, crafts and design. It is one of 10 courses covering such areas as finance for start-ups, creating a business plan, and developing “competitive intelligence”. The courses belong to the Stockholm School of Entrepreneurship, started four years ago.

“We would not like to encourage academics to leave academia, as a rule,” Sundberg says of his course. “We try to bring students out of science for a limited amount of time, to learn about patents and so forth.”

Top executives lecture on such areas as finance, staffing, negotiations and intellectual property. Finance management postgraduate student Kris Knutson at the Stockholm School of Economics was impressed by several lecturers. “I think they enjoyed what they were doing, and positive attitudes are always contagious.”

Students from the different schools and at different levels of their education are grouped in workshops to examine real problems from a company, which might

involve funding, technology or human resources. Executives of the company analyse students’ reports. “The students get to talk to the people who did work, and still work, in the companies,” Sundberg says. “They get really attentive listeners.”

Linn Hjortsberg, a Karolinska student of biomedicine, was wary of the workshop at first. “I suspected the business students would have to do all the work,” she says, “but the medical and technological knowledge of the rest of us proved to be very useful.”

Sundberg’s school has had preliminary talks about an overseas alliance with the Biomedical Technology Innovation Program at Stanford University, on which 28 graduate students of medicine, engineering and business aim to develop an innovative medical device. The three-term course combines academic studies, immersion in a medical speciality, brainstorming, modelling and business planning. Already, seven patents have been filed, says Paul Yock, who co-directs the programme with Josh Makower.

The programme is a natural for Silicon Valley, where more than 200 medical-device companies account for US\$15 billion of local economy. Many of the innovations come from Stanford faculty, including Yock, who has invented several successful cardiovascular devices and founded a company. Faculty from Stanford and other universities and business people all contribute to the teaching, and the innovations are examined by an expert panel.

Engineering and business lecturers often help students to develop prototypes, whether simple conceptual designs cut from paper or, say, a bed that adjusts to a patient’s physiology. Several major projects have already been initiated from student ideas.

Meetings are planned this autumn with staff from other universities, says Yock. The theme: “How we can mutually benefit from the programmes in technology innovation being developed around the country.” He might well have added, “And around the world.” ■

Steve Bunk is a science writer based in Alexandria, Virginia.

University of Sydney Molecular Biotechnology Program

♦ www.biotech.usyd.edu.au

Cambridge-MIT Institute

♦ www.cmi.cam.ac.uk

Stockholm School of Entrepreneurship

♦ www.sses.se

Stanford Biomedical Technology Innovation Program

♦ innovation.stanford.edu

Budding entrepreneurs: Sydney’s Molecular Biotechnology Program (left), with Tony Weiss bottom right, and (right) Karolinska course directors Carl Johan Sundberg (third from right) and Anna Nilsson (far right) with students.



Alan Windle (above) and Nick Oliver are excited by the fresh approach of the Cambridge-MIT initiative.

